

# USING NOISE TO CHARACTERIZE VISION

EDITED BY: Rémy Allard, Jocelyn Faubert and Denis G. Pelli  
PUBLISHED IN: Frontiers in Psychology



# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2016 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-753-8

DOI 10.3389/978-2-88919-753-8

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

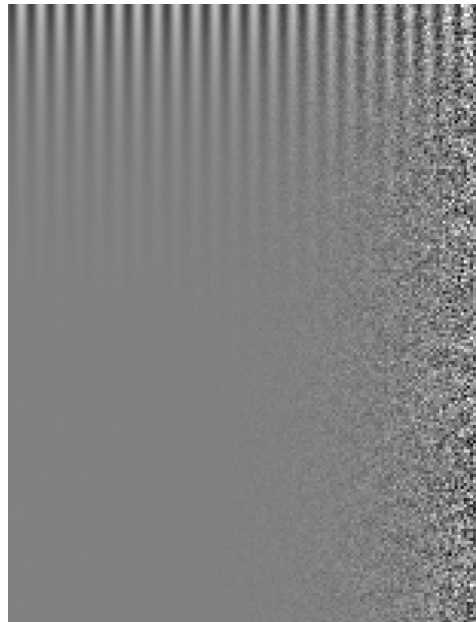
# USING NOISE TO CHARACTERIZE VISION

Topic Editors:

**Rémy Allard**, Université Pierre et Marie Curie, France

**Jocelyn Faubert**, Université de Montréal, Canada

**Denis G. Pelli**, New York University, USA



A signal grating on a white noise background. Vertically, signal contrast grows logarithmically. Horizontally, noise contrast grows logarithmically. Note that your threshold contrast (bottom edge of visibility) for the grating is flat, at the left, where the noise is invisible, and then rises with a slope of 1, to the right, indicating proportionality to noise contrast. Image by Rémy Allard.

Noise has been widely used to investigate the processing properties of various visual functions (e.g. detection, discrimination, attention, perceptual learning, averaging, crowding, face recognition), in various populations (e.g. older adults, amblyopes, migrainers, dyslexic children), using noise along various dimensions (e.g. pixel noise, orientation jitter, contrast jitter). The reason to use external noise is generally not to characterize visual processing in external noise per se, but rather to reveal how vision works in ordinary conditions when performance is limited by our intrinsic noise rather than externally added noise. For instance, reverse correlation aims at identifying the relevant information to perform a given task in noiseless conditions and measuring contrast thresholds in various noise levels can be used to understand the impact of intrinsic noise that limits sensitivity to noiseless stimuli.

Why use noise? Since Fechner named it, psychophysics has always emphasized the systematic investigation of conditions that break vision. External noise raises threshold hugely and selectively. In hearing, Fletcher used noise in his famous critical-band experiments to reveal frequency-selective channels in hearing. Critical bands have been found in vision too.

More generally, the big reliable effects of noise give important clues to how the system works. And simple models have been proposed to account for the effects of visual noise.

As noise has been more widely used, questions have been raised about the simplifying assumptions that link the processing properties in noiseless conditions to measurements in external

noise. For instance, it is usually assumed that the processing strategy (or mechanism) used to perform a task and its processing properties (e.g. filter tuning) are unaffected by the addition of external noise. Some have suggested that the processing properties could change with the addition of external noise (e.g. change in filter tuning or more lateral masking in noise), which would need to be considered before drawing conclusions about the processing properties in noiseless condition. Others have suggested that different processing properties (or mechanisms) could be solicited in low and high noise conditions, complicating the characterization of processing properties in noiseless condition based on processing properties identified in noise conditions.

The current Research Topic probes further into what the effects of visual noise tell us about vision in ordinary conditions. Our Editorial gives an overview of the articles in this special issue.

**Citation:** Allard, R., Faubert, J., Pelli, D. G., eds. (2016). Using Noise to Characterize Vision. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-753-8



# Table of Contents

- 05 Editorial: Using Noise to Characterize Vision**  
Rémy Allard, Jocelyn Faubert and Denis G. Pelli
- 08 Information processing correlates of a size-contrast illusion**  
Jason M. Gold
- 21 Phase noise reveals early category-specific modulation of the event-related potentials**  
Kornél Németh, Petra Kovács, Pál Vakli, Gyula Kovács and Márta Zimmer
- 33 Observer efficiency in free-localization tasks with correlated noise**  
Craig K. Abbey and Miguel P. Eckstein
- 46 Motion processing: The most sensitive detectors differ in temporally localized and extended noise**  
Rémy Allard and Jocelyn Faubert
- 51 Distinct mechanisms subserve location- and object-based visual attention**  
Wei-Lun Chou, Su-Ling Yeh and Chien-Chung Chen
- 59 Evidence for adjustable bandwidth orientation channels**  
Christopher P. Taylor, Patrick J. Bennett and Allison B. Sekuler
- 69 Comparison of informational vs. energetic masking effects on speechreading performance**  
Björn Lidestam, Johan Holgersson and Shahram Moradi
- 76 Effect of luminance noise on the object frequencies mediating letter identification**  
Cierra Hall, Shu Wang, Reema Bhagat and J. Jason McAnany
- 85 To characterize contrast detection, noise should be extended, not localized**  
Rémy Allard and Jocelyn Faubert
- 92 Broadband noise masks suppress neural responses to narrowband stimuli**  
Daniel H. Baker and Greta Vilidaitė
- 101 Developmental mechanisms underlying improved contrast thresholds for discriminations of orientation signals embedded in noise**  
Seong Taek Jeon, Daphne Maurer and Terri L. Lewis
- 112 Critical band masking reveals the effects of optical distortions on the channel mediating letter identification**  
Laura K. Young and Hannah E. Smithson



# Editorial: Using Noise to Characterize Vision

Remy Allard<sup>1\*</sup>, Jocelyn Faubert<sup>2</sup> and Denis G. Pelli<sup>3</sup>

<sup>1</sup> Aging in Vision and Action Laboratory, Université Pierre et Marie Curie, Paris, France, <sup>2</sup> Visual Psychophysics and Perception Laboratory, Université de Montréal, Montréal, QC, Canada, <sup>3</sup> Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

**Keywords:** noise, equivalent input noise, linear amplifier model, perceptual template model, noise image classification, bandpass noise, contrast jitter, phase noise

Auditory noise is a sound, a random variation in air pressure. More generally, random “noise” can be introduced into any stimulus, including a visual display. Noise added to the stimulus can probe the computations underlying perception of the stimulus. With power and precision, the noise, by restricting the information available, places fundamental constraints on attainable performance and processing strategy. WWII research on radar led to mathematical theorems about detectability of signals in noise, i.e., Signal Detection Theory (Peterson et al., 1954), which allow human performance to be expressed on an absolute scale of efficiency, 0–100% (Tanner and Birdsall, 1958; Pelli and Farell, 1999). Auditory noise revealed the channels of hearing in studies at Bell Labs that characterized how telephone line noise limits perception of speech (Fletcher, 1953). Studies of visual effects of photographic, x-ray, and video noise (reviewed in Pelli, 1981) led to pioneering work with artificially injected noise by Rose (1957), Stromeyer and Julesz (1972), and Solomon and Pelli (1994). Added visual noise has been widely used to characterize the computations underlying various visual tasks (e.g., detection, discrimination, letter and face recognition, search, averaging, selective attention, perceptual learning) in various populations (e.g., older adults, amblyopes, migrainers, dyslexic children). Different kinds of noise probe different aspects of the computation. For instance, spectrally filtered noise is used to determine the frequencies relevant to a given visual task (e.g., letter identification, Solomon and Pelli, 1994). Noise masking of one attribute (e.g., in luminance, color, or texture) can reveal whether another attribute is processed separately (e.g., Gegenfurtner and Kiper, 1992; Allard and Faubert, 2007, 2008). Noise image classification can reveal the visual features the observer uses to perform a visual task (e.g., Eckstein and Ahumada, 2002). Noise is also often used to characterize what limits sensitivity, such as internal noise (Pelli, 1981; Pelli and Farell, 1999; Lu and Doshier, 2008). This Research Topic issue explores effective ways to use noise to probe visual function.

“Noise” in perception experiments generally means unpredictable variation in some aspect of the stimulus. Typically, the stimulus consists of a luminance signal plus an unpredictable noise. Less often, another parameter of the signal, e.g., orientation, varies unpredictably (e.g., Dakin, 1999; Solomon, 2010; Allard and Cavanagh, 2012). Added noise is often *white*: A random sample, independent and identically distributed, is added to each pixel’s luminance. The extent of the noise is restricted, or “localized,” by a window in space and time. The spatiotemporal spectrum of the noise can be restricted by bandpass filtering to a range of orientation and frequency. Added noise that varies across space is sometimes called “pixel noise.” Most of the studies in this Research Topic issue added noise to the signal; two studies randomly jittered parameters of the signal.

In this Research Topic issue, Jeon et al. (2014) added localized white noise to investigate developmental changes in orientation discrimination through childhood. Interpreting their data

## OPEN ACCESS

### Edited and reviewed by:

Philippe G. Schyns,  
University of Glasgow, UK

### \*Correspondence:

Remy Allard  
remy.allard@inserm.fr

### Specialty section:

This article was submitted to  
Perception Science,  
a section of the journal  
Frontiers in Psychology

**Received:** 05 October 2015

**Accepted:** 22 October 2015

**Published:** 16 November 2015

### Citation:

Allard R, Faubert J and Pelli DG (2015)  
Editorial: Using Noise to Characterize  
Vision. *Front. Psychol.* 6:1707.  
doi: 10.3389/fpsyg.2015.01707

using the Perceptual Template Model (Lu and Doshier, 2008), to see how the model parameters change with age, they find that increasing age reduces internal additive noise, reduces internal multiplicative noise, and improves external noise exclusion. Using a similar noise paradigm, Chou et al. (2014) find that localized attention facilitated contrast detection due to signal enhancement, whereas object-based attention facilitated detection due to external noise exclusion. Letter identification is mediated by an octave-wide spatial frequency channel (Solomon and Pelli, 1994). Young and Smithson (2014) use spatially bandpass noise to reveal the letter identification channel in the presence of optical distortions, and find changes in the central spatial frequency of the letter-identification channel. Hall et al. (2014) find that adding white noise increased the center spatial frequency of the letter-identification channel for large but not small letters. Using pixel noises with different spectral profiles, Abbey and Eckstein (2014) find that performance approaches that of the mathematical ideal in a free-localization task (i.e., high spatial uncertainty), but is much lower in a fixed-location task (i.e., low spatial uncertainty), indicating that the human detection strategy is well-adapted to free-localization tasks. Gold (2014) use pixel noise to investigate the visual information used by the observer during a size-contrast illusion. By correlating the observers' classification decision with each pixel of the noise stimuli, they find that the spatial region used to estimate the size of the target is influenced by the size of surrounding irrelevant elements. Taylor et al. (2014) use pixel noise both as a target and a mask. The target noise is bandpass-filtered in orientation and spatial frequency, whereas the mask is white noise. They find that information used to detect the target is more optimal in the orientation domain than in the frequency domain, suggesting that observers can adjust the bandwidth of their channels in orientation, but not in spatial frequency.

Several studies examine how visual processing is affected by the extent and bandwidth of applied noise. Baker and Vilidaitė (2014) provide EEG evidence that white noise masks have a suppressive gain control effect on neural responses to grating stimuli. Happily, Allard and Faubert (2014b) note that suppressive gain control would not affect threshold

measurements in white noise. Studying motion perception, Allard and Faubert (2014a) find similar orientation and direction thresholds with and without temporally extended noise, but greater direction thresholds in temporally localized noise. This shows that the processing strategy underlying motion perception depends on the noise duration. Consistent with previous studies on contrast sensitivity (Allard and Cavanagh, 2011; Allard et al., 2013), they conclude that to measure equivalent input noise of motion processing, noise should be temporally extended (e.g., displayed continually).

Two studies randomly jittered a signal parameter. In an electrophysiological study, Németh et al. (2014) use *phase noise*, produced by randomizing phases in the Fourier domain, making the stimulus unrecognizable without affecting its spectral energy. Thus, sensitivity to phase noise suggests involvement in recognition. They find that phase-noise amplifies the P1 response to cars in the right hemisphere, but not in the left hemisphere, and that, conversely, phase-noise amplifies the P1 response to faces in the left hemisphere, but not in the right hemisphere. Lidestam et al. (2014) evaluate the effect of informational and energetic auditory noise on visual speechreading. They found that only informational auditory noise (i.e., four-talker babble) interfered with speechreading, which suggests that phonological processing is also involved in speechreading.

In sum, this Research Topic issue shows several ways to use diverse kinds of noise to probe visual processing.

## AUTHOR CONTRIBUTIONS

RA wrote the editorial, which was substantially improved by DP and proof read by JF.

## ACKNOWLEDGMENTS

Thanks to Nick Blauch, Aenne Brielmann, and Xiuyun Wu for helpful comments, and a special thanks to Najib Majaj and Manoj Raghavan. The redaction of this editorial and the organisation of this research topic were supported by ANR-Essilor SilverSight Chair and NSERC-Essilor Industrial Chair.

## REFERENCES

- Abbey, C. K., and Eckstein, M. P. (2014). Observer efficiency in free-localization tasks with correlated noise. *Front. Psychol.* 5:345. doi: 10.3389/fpsyg.2014.00345
- Allard, R., and Cavanagh, P. (2011). Crowding in a detection task: external noise triggers change in processing strategy. *Vis. Res.* 51, 408–416. doi: 10.1016/j.visres.2010.12.008
- Allard, R., and Cavanagh, P. (2012). Different processing strategies underlie voluntary averaging in low and high noise. *J. Vis.* 12:6. doi: 10.1167/12.11.6
- Allard, R., and Faubert, J. (2007). Double dissociation between first- and second-order processing. *Vis. Res.* 47, 1129–1141. doi: 10.1016/j.visres.2007.01.010
- Allard, R., and Faubert, J. (2008). First- and second-order motion mechanisms are distinct at low but common at high temporal frequencies. *J. Vis.* 8, 1–17. doi: 10.1167/8.2.12
- Allard, R., and Faubert, J. (2014a). Motion processing: the most sensitive detectors differ in temporally localized and extended noise. *Front. Psychol.* 5:426. doi: 10.3389/fpsyg.2014.00426
- Allard, R., and Faubert, J. (2014b). To characterize contrast detection, noise should be extended, not localized. *Front. Psychol.* 5:749. doi: 10.3389/fpsyg.2014.00749
- Allard, R., Renaud, J., Molinatti, S., and Faubert, J. (2013). Contrast sensitivity, healthy aging and noise. *Vis. Res.* 92, 47–52. doi: 10.1016/j.visres.2013.09.004
- Baker, D. H., and Vilidaitė, G. (2014). Broadband noise masks suppress neural responses to narrowband stimuli. *Front. Psychol.* 5:763. doi: 10.3389/fpsyg.2014.00763
- Chou, W.-L., Yeh, S.-L., and Chen, C.-C. (2014). Distinct mechanisms subserve location- and object-based visual attention. *Front. Psychol.* 5:456. doi: 10.3389/fpsyg.2014.00456
- Dakin, S. C. (1999). Orientation variance as a quantifier of structure in texture. *Spatial Vis.* 12, 1–30. doi: 10.1163/156856899X00012
- Eckstein, M. P., and Ahumada, A. J. (2002). Classification images: a tool to analyze visual strategies. *J. Vis.* 2, 1x. doi: 10.1167/2.1.i
- Fletcher, H. (1953). *Speech and Hearing in Communication*. New York, NY: Van Nostrand and Co.

- Gegenfurtner, K. R., and Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise. *J. Opt. Soc. Am. A* 9, 1880–1888. doi: 10.1364/JOSAA.9.001880
- Gold, J. M. (2014). Information processing correlates of a size-contrast illusion. *Front. Psychol.* 5:142. doi: 10.3389/fpsyg.2014.00142
- Hall, C. M., Wang, S., Bhagat, R., and McAnany, J. J. (2014). Effect of luminance noise on the object frequencies mediating letter identification. *Front. Psychol.* 5:663. doi: 10.3389/fpsyg.2014.00663
- Jeon, S. T., Maurer, D., and Lewis, T. L. (2014). Developmental mechanisms underlying improved contrast thresholds for discriminations of orientation signals embedded in noise. *Front. Psychol.* 5:977. doi: 10.3389/fpsyg.2014.00977
- Lidestam, B., Holgersson, J., and Moradi, S. (2014). Comparison of informational versus energetic masking effects on speechreading performance. *Front. Psychol.* 5:639. doi: 10.3389/fpsyg.2014.00639
- Lu, Z.-L., and Doshier, B. A. (2008). Characterizing observers using external noise and observer models: assessing internal representations with external noise. *Psychol. Rev.* 115, 44–82. doi: 10.1037/0033-295X.115.1.44
- Németh, K., Kovács, P., Vakli, P., Kovács, G., and Zimmer, M. (2014). Phase noise reveals early category-specific modulation of the event-related potentials. *Front. Psychol.* 5:367. doi: 10.3389/fpsyg.2014.00367
- Pelli, D. G. (1981). *The Effects of Visual Noise*. Department of Physiology. Cambridge: Cambridge University. Available online at: <https://psych.nyu.edu/pelli/pubs/pelli1981thesis.pdf>
- Pelli, D. G., and Farell, B. (1999). Why use noise? *J. Opt. Soc. Am. A* 16, 647–653. doi: 10.1364/JOSAA.16.000647
- Peterson, W. W., Birdsall, T. G., and Fox, W. C. (1954). The theory of signal detectability. *Trans. IRE Prof. Group Inf. Theory* 4, 171–212. doi: 10.1109/TIT.1954.1057460
- Rose, A. (1957). Quantum effects in human vision. *Adv. Biol. Med. Phys.* 2, 211–242. doi: 10.1016/B978-1-4832-3111-2.50009-2
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *J. Vis.* 10, 1–16. doi: 10.1167/10.14.19
- Solomon, J. A., and Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature* 369, 395–397. doi: 10.1038/369395a0
- Stromeyer, I. I. I., C. F., and Julesz, B. (1972). Spatial-frequency masking in vision: critical bands and spread of masking. *J. Opt. Soc. Am.* 62, 1221–1232. doi: 10.1364/JOSA.62.001221
- Tanner, W. P. Jr., and Birdsall, T. G. (1958). Definitions of  $d'$  and  $\eta$  as psychophysical measures. *J. Acoust. Soc. Am.* 30, 922–928. doi: 10.1121/1.1909408
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2014). Evidence for adjustable bandwidth orientation channels. *Front. Psychol.* 5:578. doi: 10.3389/fpsyg.2014.00578
- Young, L. K., and Smithson, H. E. (2014). Critical band masking reveals the effects of optical distortions on the channel mediating letter identification. *Front. Psychol.* 5:1060. doi: 10.3389/fpsyg.2014.01060

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Allard, Faubert and Pelli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Information processing correlates of a size-contrast illusion

Jason M. Gold\*

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Andrew M. Haun, Harvard Medical School, USA

Craig Abbey, University of California at Santa Barbara, USA

## \*Correspondence:

Jason M. Gold, Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, IN 47405, USA  
e-mail: jgold@indiana.edu

Perception is often influenced by context. A well-known class of perceptual context effects is perceptual contrast illusions, in which proximate stimulus regions interact to alter the perception of various stimulus attributes, such as perceived brightness, color and size. Although the phenomenal reality of contrast effects is well documented, in many cases the connection between these illusions and how information is processed by perceptual systems is not well understood. Here, we use noise as a tool to explore the information processing correlates of one such contrast effect: the Ebbinghaus–Titchener size-contrast illusion. In this illusion, the perceived size of a central dot is significantly altered by the sizes of a set of surrounding dots, such that the presence of larger surrounding dots tends to reduce the perceived size of the central dot (and *vice versa*). In our experiments, we first replicated previous results that have demonstrated the subjective reality of the Ebbinghaus–Titchener illusion. We then used visual noise in a detection task to probe the manner in which observers processed information when experiencing the illusion. By correlating the noise with observers' classification decisions, we found that the sizes of the surrounding contextual elements had a direct influence on the relative weight observers assigned to regions within and surrounding the central element. Specifically, observers assigned relatively more weight to the surrounding region and less weight to the central region in the presence of smaller surrounding contextual elements. These results offer new insights into the connection between the subjective experience of size-contrast illusions and their associated information processing correlates.

**Keywords:** visual illusion, response classification, noise, efficiency, ideal observer

## INTRODUCTION

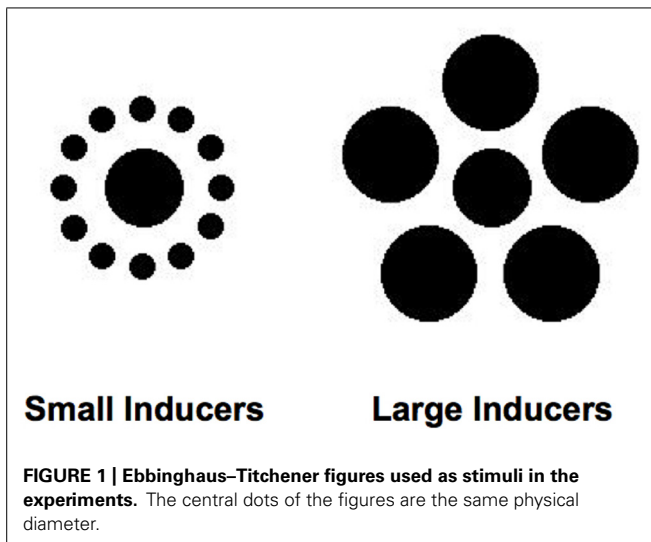
Context can often exert a significant influence on perception. Famous examples of context effects include crowding (Bouma, 1970), word superiority effects (Johnston and McClelland, 1974), configural superiority effects (Pomerantz et al., 1977), the kinetic depth effect (Wallach and O'Connell, 1953), point-light biological motion perception (Johansson, 1973), Gestalt grouping and perceptual organization (Koffka, 1935), and visual completion (Kanizsa, 1979). Another related category of context effects involves the perceptual consequences of introducing contrast between elements within a display. Examples of contrast effects include lightness and brightness contrast illusions (Cornsweet, 1970; Adelson, 1993; Gilchrist et al., 1999), color contrast illusions (Jameson and Hurvich, 1964; Lotto and Purves, 2000), and size-contrast illusions (Coren and Girgus, 1978).

In the cases of lightness, brightness, and color contrast illusions, the underlying physiological and information processing mechanisms that mediate these effects have been studied extensively (e.g., Jameson and Hurvich, 1964; Cornsweet, 1970; Adelson, 1993; Lotto and Purves, 2000). In the case of size-contrast illusions, most research has focused on exploring the conditions that are most favorable for inducing the illusions (e.g., Girgus et al., 1972; Coren and Girgus, 1978; Jaeger, 1978; Weintraub, 1979; Weintraub and Schneck, 1986; Rose and Bressan, 2002; Roberts et al., 2005; Daneyko et al., 2011), demonstrating the behavioral impact of the illusions in various tasks (e.g., Jaeger,

1978; Pavlova and Sokolov, 2000; Haffenden et al., 2001; Rose and Bressan, 2002; Westwood and Goodale, 2003; Handlovsky et al., 2004; Muller and Busch, 2006; Im and Chong, 2009; Sperandio et al., 2010, 2012), or using the illusions as research tools to understand various aspects of perceptual processing, such as whether apparent size is coded in pre-attentive vision (Busch and Muller, 2004) and whether there are two separate visual processing streams (e.g., Aglioti et al., 1995; Milner and Goodale, 1995; Goodale and Humphrey, 1998).

One size-contrast illusion, the Ebbinghaus–Titchener illusion (Titchener, 1901), has been used most extensively in this research. **Figure 1** shows the canonical form of the Ebbinghaus–Titchener illusion. When most observers view these figures, the central dot is judged to be significantly larger when encircled by smaller dots (left side of **Figure 1**) than when surrounded by larger dots (right side of **Figure 1**). The magnitude of this effect has been shown to depend upon many additional factors, including the distance between the central dot and the surrounding dots, the number and density of surrounding dots, the similarity between the central and surrounding dots, and even the age, sex, and culture of the observer (Massaro and Anderson, 1971; Coren and Girgus, 1978; Weintraub, 1979; Weintraub and Schneck, 1986; Choplin and Medin, 1999; Phillips et al., 2004; Roberts et al., 2005; de Fockert et al., 2007; Daneyko et al., 2011). Nevertheless, the subjective experience of the Ebbinghaus–Titchener illusion is quite reliable and robust for most observers under a wide range of conditions.





Despite the extensive amount of research that has involved this size-contrast effect, the connection between the subjective experience of the illusion and the specific manner in which information is processed by the visual system is not well understood. There are many possible ways in which the experience of the illusion might map on to how observers make use of information when performing tasks that rely upon the part of the stimulus that is perceptually altered by the presence of the inducing elements. For example, observers might make use of a relatively larger region of the central portion of the stimulus in the presence of smaller inducing elements. Another possibility is that observers might differentially rely upon the regions within and immediately surrounding the central dot, depending upon the size of the inducing elements. Alternatively, there may be no link or connection between observers' subjective experience of the illusion and how they make use of information in tasks involving these stimuli.

Thus, the goal of the current study was to directly address this question by exploring the underlying information processing correlates associated with the perception of the Ebbinghaus-Titchener size-contrast illusion in a perceptual task. We approached this problem by first measuring and verifying the traditional subjective size-contrast effects associated with the illusion. We then employed these same stimuli to be used within the context of a performance-based rather than a subjective judgment task. Specifically, we had observers perform a simple detection task with the central dot of Ebbinghaus-Titchener figures under conditions of varying context (i.e., in the presence of larger or smaller surrounding dots). We chose a detection task as a starting point because of its relative simplicity. Observers performed this task with stimuli that were embedded in high contrast pixel noise, which allowed us to measure the impact of context on two related aspects of information processing: (a) the overall *efficiency* with which observers make use of information (i.e., their performance relative to a statistically optimal or *ideal observer*); and (b) the perceptual strategy or "template" used by observers, determined by correlating the noise shown across trials with observers' decisions

(i.e., *response classification*). A similar approach has been used successfully to explore the information processing correlates associated with brightness-contrast context effects (Shimozaki et al., 2005).

## MATERIALS AND METHODS

### PARTICIPANTS

Three observers (two males, mean age 20) participated in both experiments. All were paid for their participation, gave written consent and had normal or corrected-to-normal visual acuity (self-reported). Two were naïve to the purposes of the experiments and one was a paid laboratory research assistant (observer PM). The study was approved by the Indiana University Human Research Protection Program.

### APPARATUS

All stimuli were displayed on a Sony Trinitron G520 CRT monitor (resolution: 1024 pixels × 768 pixels; size: 38.25 cm × 28.5 cm; refresh rate: 85 Hz). The display was calibrated using a Minolta LS-100 photometer. The background was fixed at a luminance of 85 cd/m<sup>2</sup>, and the CRT provided the only source of illumination during the experiment. Viewing distance was fixed at 130 cm using a head/chin rest. All aspects of the experiment, including stimulus generation, presentation, and data analysis, were carried out within the MATLAB programming environment (version 7.1) using in-house software and the extensions provided by the psychophysics toolbox (Brainard, 1997).

### STIMULI

Stimuli consisted of a central dot (45 pixels in diameter, 0.74°) surrounded by a series of "inducing" dots (Figure 1). In the Small Inducers condition, there were 12 surrounding dots of equal size (15 pixels in diameter, 0.25°), equidistant from the central dot (45 pixels from the midpoint of each inducer to the midpoint of the central dot, 0.74°) and equally spaced around the perimeter of a virtual circle centered upon the central dot. In the Large Inducers condition, there were five dots of equal size (55 pixels in diameter, 0.9°), equidistant from the central dot (60 pixels from the midpoint of each inducer to the midpoint of the central dot, 0.98°) and equally spaced around the perimeter of a virtual circle centered upon the central dot. In Experiment 1, a central dot of variable size with no surrounding inducing elements was also used to obtain estimates of perceived size.

All stimuli were defined in terms of contrast, with the contrast at each pixel defined as the luminance value relative to the background luminance (i.e.,  $L_{\text{pixel}} - L_{\text{background}} / L_{\text{background}}$ ). Stimuli were negative in contrast (i.e., darker than the background). In Experiment 1, each pixel of the entire stimulus was set to the maximum displayable negative contrast value (−0.87). In Experiment 2, only the pixels in the inducing dots were set to the maximum displayable contrast value. For the remaining image pixels, the contrast energy was manipulated across trials using a 2-down, 1-up adaptive staircase procedure in order to maintain constant performance, as well as obtain contrast energy detection thresholds. Contrast energy is defined as the sum of the squared pixel contrast values multiplied by the area of an individual pixel, i.e.:

$$E = \sum_{i=1}^n C_i^2 a, \quad (1)$$

where  $n$  is the number of image pixels,  $C$  is the contrast at each pixel, and  $a$  is the area of an individual pixel, expressed in degrees squared (Tjan et al., 1995).

In addition, Gaussian white contrast noise of a fixed variance ( $\sigma^2 = 0.16$ ,  $\text{NSD} = 2.7e^{-4}$ ) was added to all pixels (except for the inducing dots) within a 200 pixel  $\times$  200 pixel ( $3.27^\circ \times 3.27^\circ$ ) region centered at the central dot. Noise samples that exceeded  $\pm 2$  standard deviations were discarded and replaced with fresh samples. This insured that the noise distribution retained its normal shape while removing any values that exceeded the maximum displayable positive and negative contrast values. The stimulus duration was 43 frames ( $\sim 500$  ms).

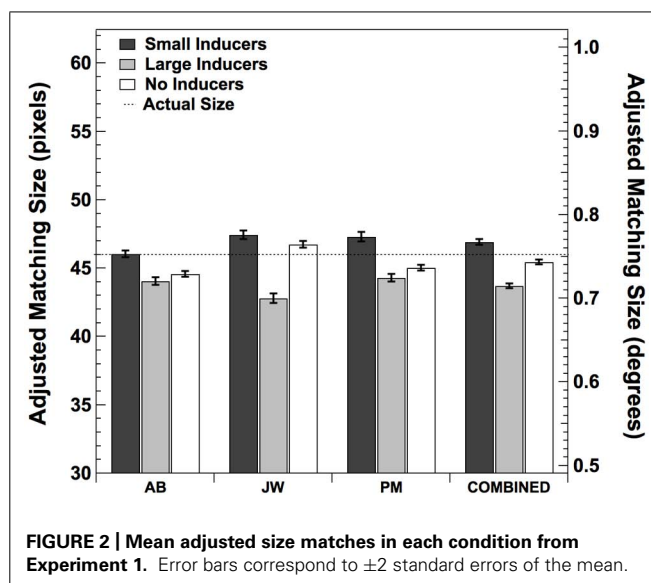
### THRESHOLD ESTIMATION

Contrast energy detection thresholds in Experiment 2 were estimated by fitting Weibull psychometric functions to the staircase data in each condition and interpolating to find the contrast energy value that corresponded to 71% correct performance. Bootstrap simulations (Efron and Tibshirani, 1993) were carried out in order to estimate the error associated with each threshold estimate (500 simulated experiments per threshold).

### PROCEDURE

In Experiment 1, on each trial either the Small Inducers stimulus, Large Inducers stimulus, or a single isolated central dot stimulus with no surrounding inducers (No Inducers) was displayed in the center of the CRT (all were noise-free and set to the maximum displayable negative contrast). A second isolated dot figure (also set to the maximum displayable negative contrast) simultaneously appeared on the display and was offset 200 pixels ( $3.27^\circ$ ) to the right and 200 pixels down from the central stimulus. On half of the trials, the size of the offset dot was initially set at 15 pixels ( $0.25^\circ$ ); on the other half of the trials, the size of the offset dot was initially set at 68 pixels ( $1.11^\circ$ ; chosen randomly on each trial with equal probability). Once the stimuli were displayed, the observer was instructed to use two keys to manipulate the size of the offset dot so that it appeared to match the size of the central dot. Three observers completed 20 trials in each stimulus condition (i.e., Small Inducers, Large Inducers, and No Inducers). Trials were blocked by condition, with one observer completing each of the three conditions first.

In Experiment 2, either the Small Inducers or Large Inducers stimulus was displayed in the center of the CRT (a No Inducers condition was not included due to the complicating effects of spatial uncertainty at low contrast in the absence of inducers). The stimuli were shown in high contrast noise, and the contrast energy of the central dot was varied across trials to keep performance at roughly 71% correct throughout the experiment. On half of the trials, the central dot was actually present; on the remaining half of the trials, the dot was absent (randomly chosen). The observer's task was to indicate whether or not the central dot had been present on a given trial. Accuracy feedback was given in the



form of a high or low beep. Each observer from Experiment 1 participated in 10,000 trials in both stimulus conditions, measured over the course of approximately 3 weeks. Trials were blocked by condition, with two observers completing the Large Inducers condition first and the other observer completing the Small Inducers condition first.

## RESULTS

### EXPERIMENT 1: SUBJECTIVE RATINGS

The purpose of Experiment 1 was to verify the presence and measure the magnitude of the subjective size-contrast illusion produced by the Ebbinghaus–Titchener patterns shown in Figure 1. Three observers repeatedly adjusted an isolated circle to match the perceived size of the central dot in each stimulus condition. The mean adjusted matching sizes for each observer as well as the mean values across observers are shown in Figure 2. These data show there was a consistent effect of the presence of the inducers, with Large Inducers producing smaller estimates than Small Inducers and No Inducers falling in between. A one-way repeated measures ANOVA revealed a significant effect of condition [ $F(2,2) = 7.74$ ,  $p < 0.05$ ]. *Post hoc* comparisons using the Tukey HSD test indicated that the mean estimate in the Small Inducers condition was significantly greater than the mean estimate in the Large Inducers condition ( $p < 0.05$ ). There were no significant differences between the mean estimates in the No Inducers condition and either the Small or Large Inducers conditions. Thus, Experiment 1 established that our stimuli produced significant size-contrast illusions for all three of our observers.

### EXPERIMENT 2: BEHAVIORAL PERFORMANCE, EFFICIENCY AND CLASSIFICATION IMAGES

Experiment 2 was designed to explore what impact the Ebbinghaus–Titchener size-contrast illusion has on behavioral performance, the efficiency of information use and observers' classification strategies when they are asked to perform a task that directly relies on the features that are perceptually distorted by the



illusion. We asked the same three observers that participated in Experiment 1 to perform a detection task, in which the contrast of the central dot of the Ebbinghaus–Titchener figure was varied across trials in order to measure contrast detection thresholds in each condition. The stimuli were shown in high contrast Gaussian white noise (with the exception of the locations where the inducers appeared, which were always noise-free and shown at the maximum displayable negative contrast).

Detection thresholds for all three human observers as well as the mean values across observers are shown in **Figure 3A**. The performance of a statistically optimal or “ideal observer” was also measured in each condition (Green and Swets, 1966; Braje et al., 1995). Such an observer uses a decision rule that maximizes the posterior probability of choosing whether or not the central dot was present (see Braje et al., 1995 for a detailed description the ideal decision rule in a detection task). The ideal observer’s thresholds were estimated by carrying out Monte Carlo simulations in each condition for the same number of trials as the human observers (10,000). The ideal observer’s thresholds are plotted in the leftmost side of the **Figure 3A**. Finally, the ratio of ideal to human threshold (*efficiency*) was computed for each human observer in each condition (**Figure 3B**). As expected, the ideal observer’s thresholds were the same for the Large and Small Inducers conditions. Although human thresholds differed by about an order of magnitude from those of the ideal observer (yielding efficiencies of  $\sim 10\%$ ), there was no discernable effect of inducer condition on human efficiency. A two-tailed paired-samples *t*-test confirmed that the effect of condition for the human observers was not statistically significant;  $t(2) = -0.68$ ,  $p = 0.57$ .

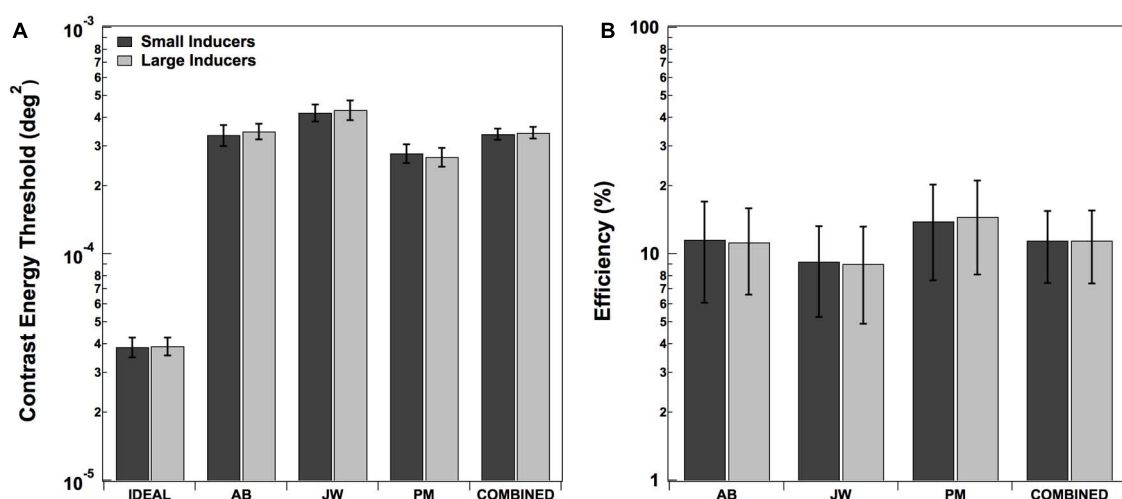
In addition to the thresholds and efficiencies, we used the noise presented over the course of the experiment to generate classification images for each observer in each condition (Ahumada and Lovell, 1971; Ahumada, 2002; Murray et al., 2002). Classification images were computed by first sorting the noise for a given

observer in a given condition according to the Stimulus (*present*, *absent*)–Response (*present*, *absent*) combination. Next, the noise was averaged within each stimulus–response (S–R) pairing and then combined to form a single *classification image C*:

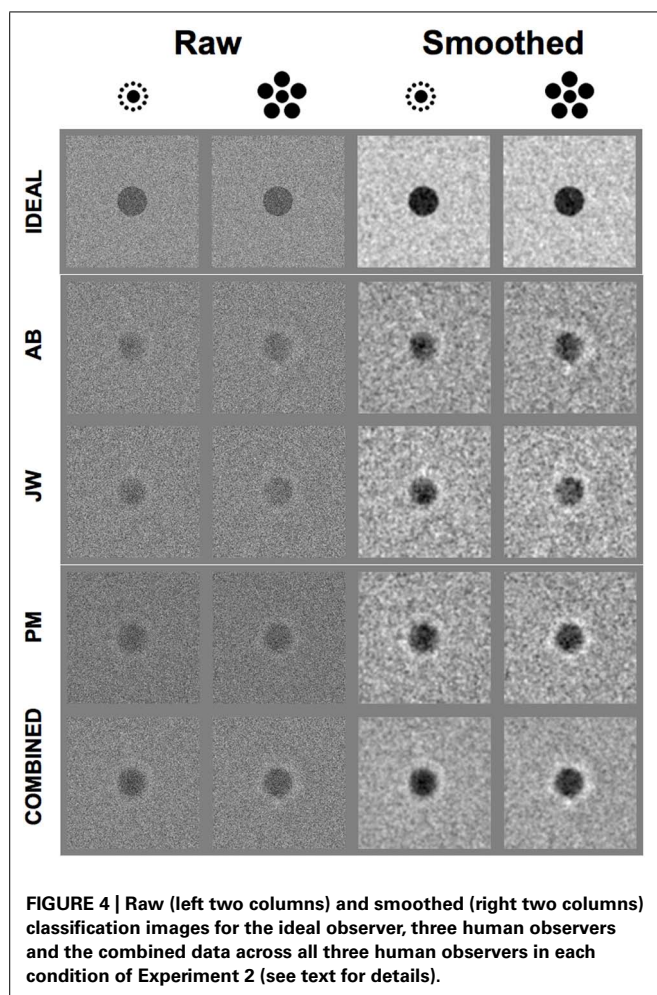
$$C = (S_{\text{absent}}R_{\text{present}} + S_{\text{present}}R_{\text{present}}) - (S_{\text{absent}}R_{\text{absent}} + S_{\text{present}}R_{\text{absent}}) \quad (2)$$

The resulting classification images show the relative weight assigned to each pixel in the display by the observer over the course of the experiment. The classification images in each condition for each human observer as well as the ideal observer are shown in the left two columns of **Figure 4**. The bottom row of these columns also shows the classification images generated by combining all of the trials across the three human subjects in each condition. The right two columns of **Figure 4** show the same classification images smoothed by a small ( $7 \text{ pixel} \times 7 \text{ pixel}$ ,  $0.11^\circ \times 0.11^\circ$ ) convolution kernel. Note that the regions where the inducing elements appeared are not noise-free. These regions were simply populated by random noise samples when computing the classification images. This was done in order to avoid inducing the illusion itself when visualizing the data. That is, presenting the classification images with the inducing element regions set to some constant value (e.g., 0), would potentially affect the perceived size of the central regions, and thus make it difficult to visually compare them across conditions. Adding random noise samples to these regions when generating the classification images allows them to blend naturally into their neighboring background regions.

These data show that the human observers adopted a very specific strategy in both conditions. Namely, each human observer evaluated the contrast of both the inner region (where the central dot appeared) as well as a circular region that surrounded the central dot. In addition, observers responded differentially to contrast in these two regions. Specifically, if the contrast of



**FIGURE 3 | Contrast energy thresholds (A) and efficiencies (B) in each condition of Experiment 2.** Error bars for individual observers correspond to  $\pm 2$  standard deviations, estimated by bootstrap simulations. Error bars on combined thresholds correspond to  $\pm 2$  standard errors of the mean.



the noise was negative in the region of the central dot, observers were more likely to respond “present” (or, if the contrast of the noise in this region was positive, observers were more likely to respond “absent”). However, the opposite was true in the annular region that surrounded the central dot: if the contrast of the noise was positive in this region, observers were more likely to respond “present” (or, if the contrast of the noise in this region was negative, observers were more likely to respond “absent”). Note that this strategy of using an annular region surrounding the central dot is not ideal: the ideal observer uses only the central dot region where the stimulus was actually present; the surrounding region carries no physical information for performing the task. Similar center-surround effects have been reported for tasks requiring observers to detect or discriminate a centralized target in noise (e.g., Shimozaki et al., 2005).

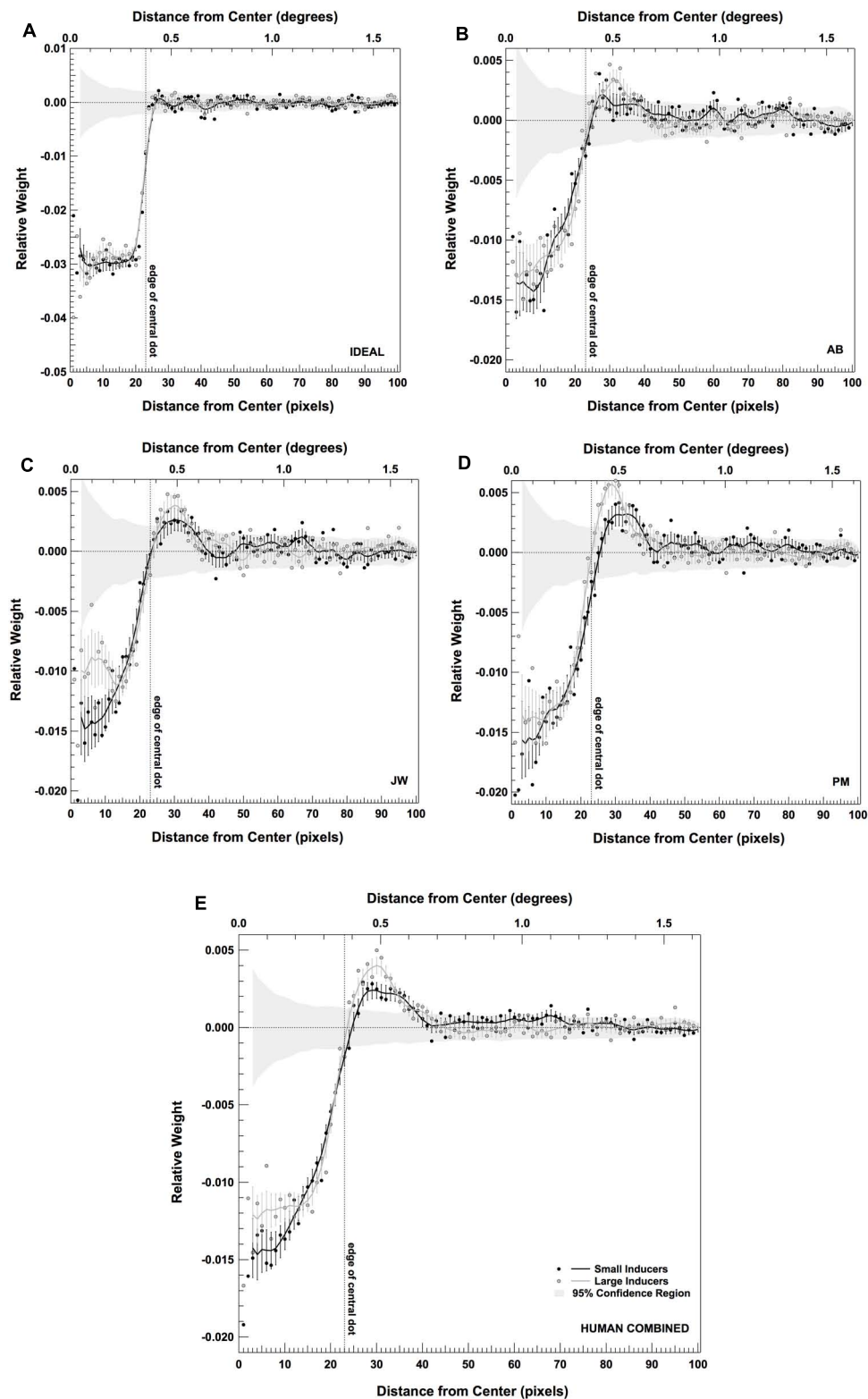
The results of the classification image analysis are consistent with the idea that, unlike the ideal observer, human observers were comparing the contrast within the region of the central dot to the contrast immediately surrounding the central dot region in order to make their classification decisions. However, this center-surround effect appears to be independent of the presence of the Large and Small Inducers. To explore the effect of inducer size more closely, we took advantage of the

circular-symmetric shape of the central portion of our stimuli and radially averaged the raw classification images (Abbey and Eckstein, 2002, 2007). This produced a set of one-dimensional classification images that revealed the weights observers assigned to each distance from the midpoint of the central dot in each condition.

The results of this radial classification image analysis are shown in **Figure 5**. **Figures 5A–D** plots the results for an individual observer in each condition (including the ideal observer; **Figure 5A**). **Figure 5E** plots the results when the data are combined across all three human observers. Individual points in each plot correspond to the raw classification image weights. The solid lines correspond to the average classification image generated by running 500 bootstrap simulations (generated by sampling the data in each condition with replacement for each observer) and then smoothing these images with a convolution kernel. The error bars on each smoothed curve correspond to  $\pm 2$  standard deviations, calculated from the bootstrap simulations. Finally, the dashed vertical line in each plot shows the location of the edge of the central dot. These data reveal that, although the spatial extent of the regions used by human observers was similar across conditions, the relative weights assigned to the central and the surrounding regions were markedly different. Specifically, all three human observers tended to place relatively more weight upon the central dot region in the presence of Small Inducers and relatively more weight on the surrounding region in the presence of Large Inducers.

We ran two sets of statistical analyses in order to verify these effects. The first was a parametric test for the overall statistical significance of (a) the difference between each raw radial classification image and the null hypothesis of zero correlation; and (b) the difference between the raw radial classification images obtained in the presence of Small vs. Large Inducers for each observer and the data combined across observers. We used the single-sample Hotelling  $T^2$  statistic to test against the null hypothesis of zero correlation and the independent two-sample Hotelling  $T^2$  statistic to test for significant differences between inducer conditions (for details on computing Hotelling  $T^2$  statistics, see Abbey and Eckstein, 2002; Eckstein et al., 2002; Shimozaki et al., 2005). The results of these tests are shown in **Table 1** (single-sample tests) and **Table 2** (two-sample tests). These data confirm that the overall classification images for all observers in both conditions significantly differed from a zero-correlation classification image, and that the overall difference between the Small and Large Inducer classification images was highly significant for all observers.

We next gauged the likelihood that the weights at each location deviated significantly from what would be expected purely by chance by generating a series of classification images that were created by randomly choosing noise images on each trial of the experiment. Specifically, these classification images were created by replacing the noise samples generated in our experiment with newly generated noise samples and re-computing the classification images. We generated 200 of these random classification images for the individual subject data sets (10,000 trials) and another 200 for the collapsed data set (30,000 trials). We then computed the mean and standard deviation across these replications in order to



**FIGURE 5 |** Radially averaged classification images for the ideal observer (A), three human observers (B–D) and the combined data across all three human observers (E) in each condition of Experiment 2. Error bars correspond to  $\pm 2$  standard deviations,

estimated by bootstrap simulations. The gray band shows a region  $\pm 2$  standard deviations around what would be expected from a purely random classification image for the same number of trials (estimated by bootstrap simulations; see text for details).

generate the gray band shown in each panel of **Figure 5**. Thus, this band represents  $\pm 2$  standard deviations around the mean randomly generated classification image. These simulations show that the human classification image weights within and directly surrounding the central dot fell well outside of this region (with the exception of the locations corresponding to the border between the two regions).

In addition to this spatial classification image analysis, we also explored the effects of inducer size on observer's use of information across spatial frequencies. Specifically, we transformed each of the classification images shown in the left two columns of **Figure 4** into the spatial frequency domain, and computed the average squared amplitude at each spatial frequency in each image (**Figure 6**). As in **Figure 5**, **Figures 6A–D** plots the results for an individual observer in each condition (including the ideal observer; **Figure 6A**). **Figure 6E** plots the results when the data are combined across all three human observers. Individual points in each plot correspond to the average squared amplitude in the classification image at a particular spatial frequency. The error bars on each point correspond to  $\pm 2$  standard deviations, computed by running 500 bootstrap simulations (generated by sampling

the data in each condition with replacement for each observer). These data reveal that observers adopted a strategy that involved placing relatively more weight on slightly higher frequencies in the presence of Large Inducers (peak at  $\sim 6$  c/deg in the presence of Small Inducers and  $\sim 9$  c/deg in the presence of Large Inducers).

## DISCUSSION

The goal of our experiments was to explore the information processing correlates of the Ebbinghaus–Titchener size-contrast illusion. In Experiment 1, we replicated the results of many previous experiments by demonstrating the subjective reality of this illusion. In Experiment 2, we asked observers perform a detection task with the same stimuli used in Experiment 1, albeit embedded in high contrast visual noise. By comparing observers' contrast detection thresholds in this task to that of an ideal observer, we found that the efficiency with which observers used information did not depend upon the size of the inducing elements. By computing the correlation between the noise contrast at each pixel and the observers' responses across trials, we found that observers tended to place relatively more weight upon the region surrounding the inner dot in the presence of Large Inducers and relatively more weight upon the region inside the inner dot in the presence of Small Inducers. We also found that observers tended to place relatively more weight upon slightly higher frequencies in the presence of Large Inducers (i.e.,  $\sim 9$  c/deg) and relatively more weight upon slightly lower frequencies in the presence of Small Inducers (i.e.,  $\sim 6$  c/deg).

So how do we interpret these findings? First, consider the finding that efficiency was unaffected by the size of the inducing elements. On the one hand, the subjective ratings given by observers in Experiment 1 showed that observers' judgments of size are farther from veridical in the presence of Large than Small Inducers. In addition, the tendency of human observers to assign relatively greater weight to the center and relatively less weight to the surround in the presence of Small Inducers is more similar to the weights used by the ideal observer, which would predict efficiency should be greater in the presence of Small than Large Inducers (Murray et al., 2005). However, there are several reasons why we might not expect to see such variations in efficiency across conditions in Experiment 2. First, there is no necessary relationship between an observer's subjective experience of an illusion and their ability to perform a task with the stimuli that produce the illusion. That is, it is unclear how the misjudgments in perceived size found in Experiment 1 should map on to an observer's ability to make use of information in Experiment 2. The most we can ultimately hope for is that there may be some correlation between the two (Teller, 1984). Second, the task we asked observers perform does not directly rely upon the precision of size judgments, only the ability to detect the presence of the central dot. As such, it is unclear that greater misjudgments in size would negatively affect performance in such a task. And finally, the prediction that greater similarity between the human and ideal classification images should lead to greater efficiency assumes a number of other factors known to affect efficiency are invariant across conditions (e.g., internal noise, point-wise non-linearities; Murray et al., 2005). More detailed measurements and analyses than those

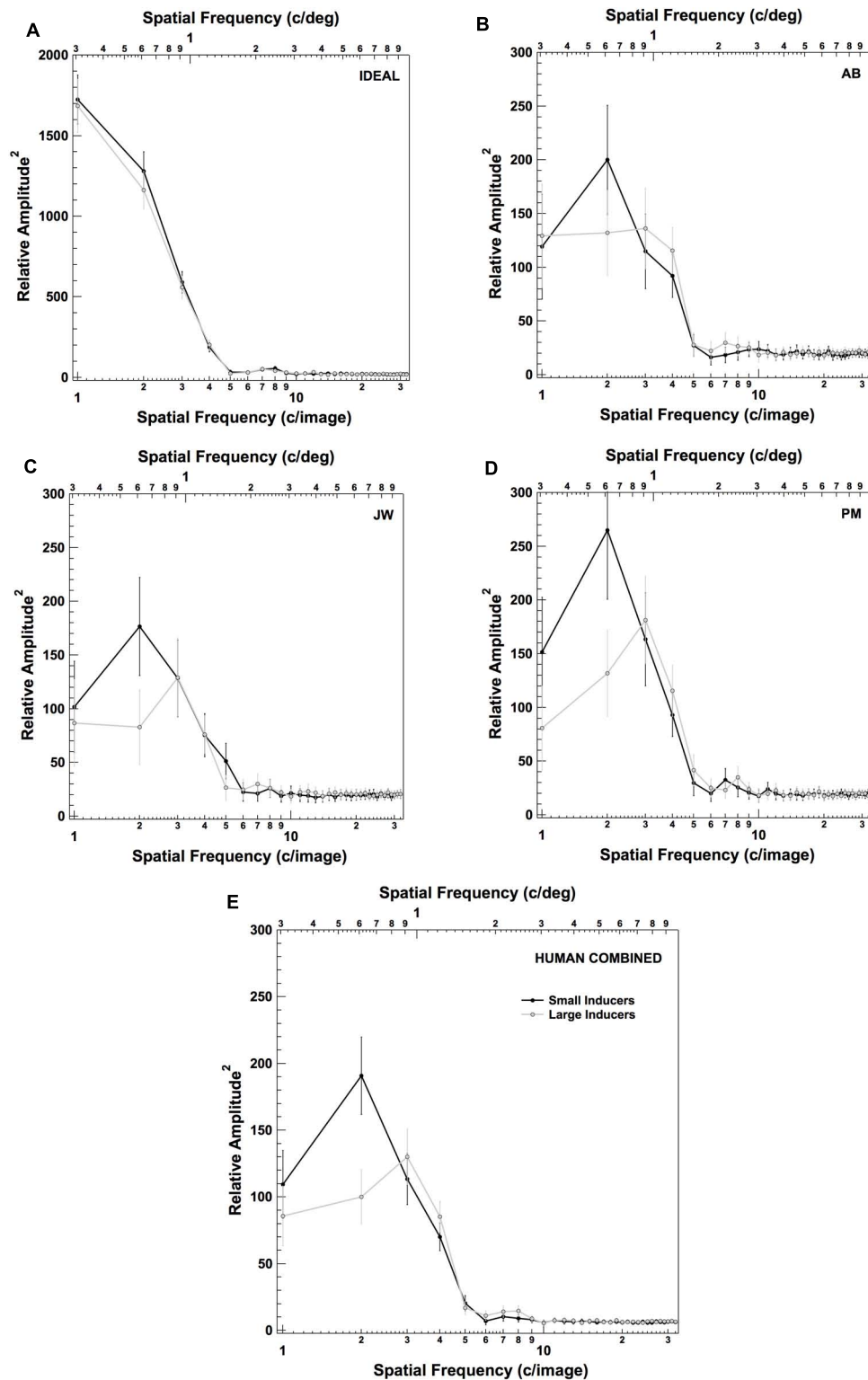
**Table 1 | Degrees of freedom,  $F$  values and  $p$  values obtained from the single-sample Hotelling  $T^2$  statistic, testing the radial classification images obtained for each human observer and the combined data across observers in each condition against the null hypothesis of zero correlation.**

Observer	df numerator	df denominator	$F$ value	$p$ Value
<b>Small Inducers vs. zero</b>				
AD	99	9901	149.48	<0.0001
JW	99	9901	159.84	<0.0001
PM	99	9901	207.19	<0.0001
COMBINED	99	29901	459.71	<0.0001
<b>Large Inducers vs. zero</b>				
AD	99	9901	158.28	<0.0001
JW	99	9901	137.10	<0.0001
PM	99	9901	201.94	<0.0001
COMBINED	99	29901	455.02	<0.0001

**Table 2 | Degrees of freedom,  $F$  values and  $p$  values obtained from the independent two-sample Hotelling  $T^2$  statistic, testing for the difference between the radial classification images obtained for each human observer and the combined data across observers with Large and Small Inducers.**

<b>Large Inducers vs. Small Inducers</b>				
Observer	df numerator	df denominator	$F$ value	$p$ Value
AD	99	19900	23.54	<0.0001
JW	99	19900	27.13	<0.0001
PM	99	19900	29.72	<0.0001
COMBINED	99	59900	33.21	<0.0001





**FIGURE 6 |** Frequency domain representation of the raw classification images for the ideal observer (A), three human observers (B–D) and the combined data across all three human observers (E) in each condition of Experiment 2. Error bars correspond to  $\pm 2$  standard deviations (estimated by bootstrap simulations; see text for details).

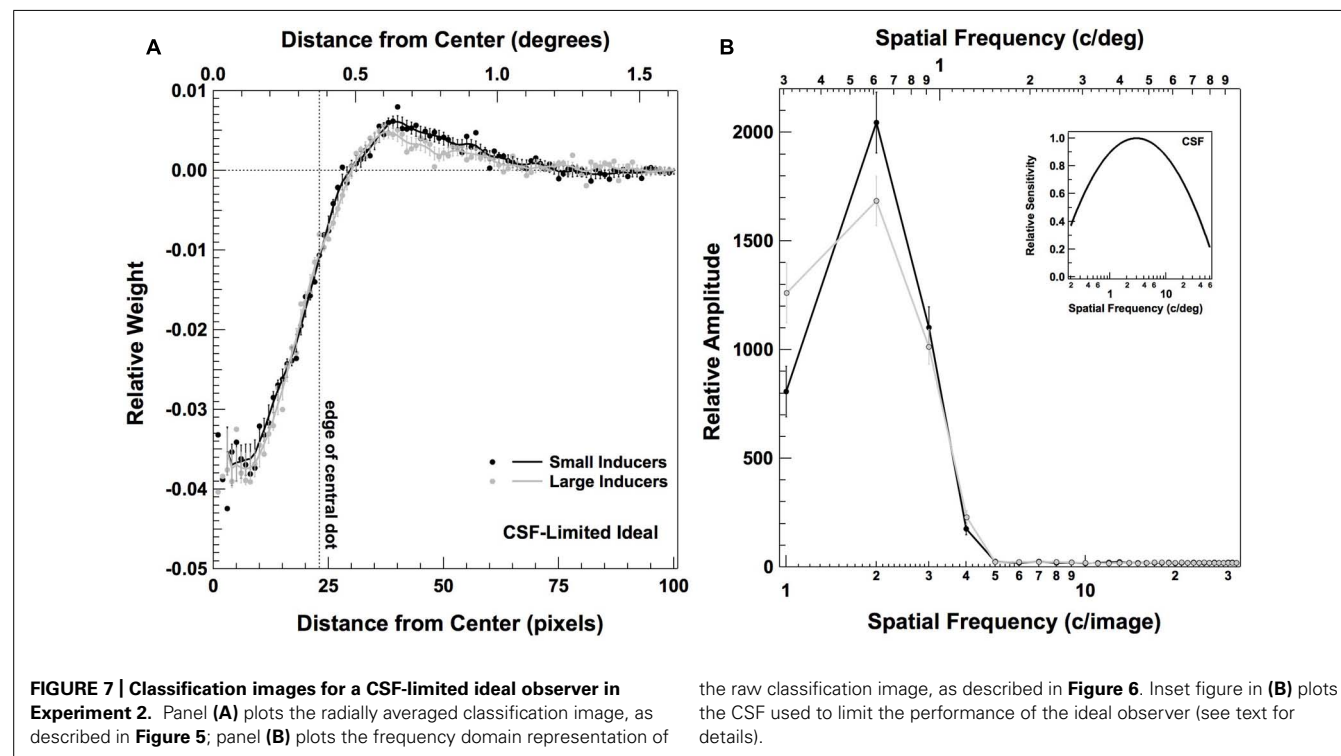
reported here would be required in order to properly test this prediction.

Despite the equivocal nature of the efficiencies obtained in Experiment 2, observers nevertheless exhibited the use of a markedly different strategy in the presence of Large and Small Inducers. So why observers might have adopted such different strategies within different contexts? One potential source of this effect could be the spatial frequency filtering that takes place during the early stages of visual processing (Geisler, 1989). We explored this possibility by building an ideal observer that was ideal in all respects, with the exception that it was limited by the foveal contrast sensitivity function (CSF) of a normal adult human (inset of **Figure 7B**). The CSF was generated from the fits reported in Watson (2000). The CSF-limited ideal observer analysis was carried out in a fashion similar to that described by Chung et al. (2002) and Nandy and Tjan (2008). Specifically, the CSF was applied to both the noise-free signals (with the inducing elements present) as well as the noise-free templates (without the inducing elements present) in the frequency domain in each condition. On each trial, unfiltered white noise of the same variance as used in the original experiments was added to the filtered signal, and the filtered templates were used to compute the likelihoods for each alternative (i.e., present, absent). All other aspects of the CSF-limited ideal observer analysis were the same as those used for the original ideal observer analysis.

**Figure 7** shows the classification images obtained from a simulated experiment carried out with our CSF-limited ideal observer performing the same detection task and for the same number of trials as our human observers. **Figure 7A** plots the radially averaged classification image, computed in the same fashion as the

plots in **Figure 5**; **Figure 7B** shows the Fourier representation of the classification image, computed in the same fashion as the plots in **Figure 6**. Interestingly, these data reveal that the center-surround weighting in the human classification images is well predicted by the filtering characteristics of the human visual system. That is, unlike the true ideal observer, our human observers and the CSF-limited ideal observer both give weight to the area directly surrounding the central dot as well as the area within the central dot. Despite these similarities, there appear to be no discernable differences in the weighting of the center relative to the surround in the presence of Large vs. Small Inducers for the CSF-limited ideal observer. We also do not see the characteristic shift toward weighting slightly higher spatial frequencies in the presence of Large relative to Small Inducers that we found with our human observers. Thus, although the human CSF accurately predicts the gross center-surround characteristics of the human observers' classification images, the results of our simulation suggest it is unlikely that the human observers' tendency to differentially weight the center and surround in the presence of different sized inducers was due to the spatial frequency filtering that takes place during the early stages of visual processing. The connection between the variations in perceived size of the central element and the differential weighting of the center and surround thus remains unclear.

Of course, it is always possible that the magnitude of the Ebbinghaus–Titchener size-contrast illusion is greatly reduced or even non-existent when the central dot is presented at low contrast in large amounts of pixel noise, as it was in our experiments. One argument against this idea is that fact that our response classification analyses showed that there were significant differences in how



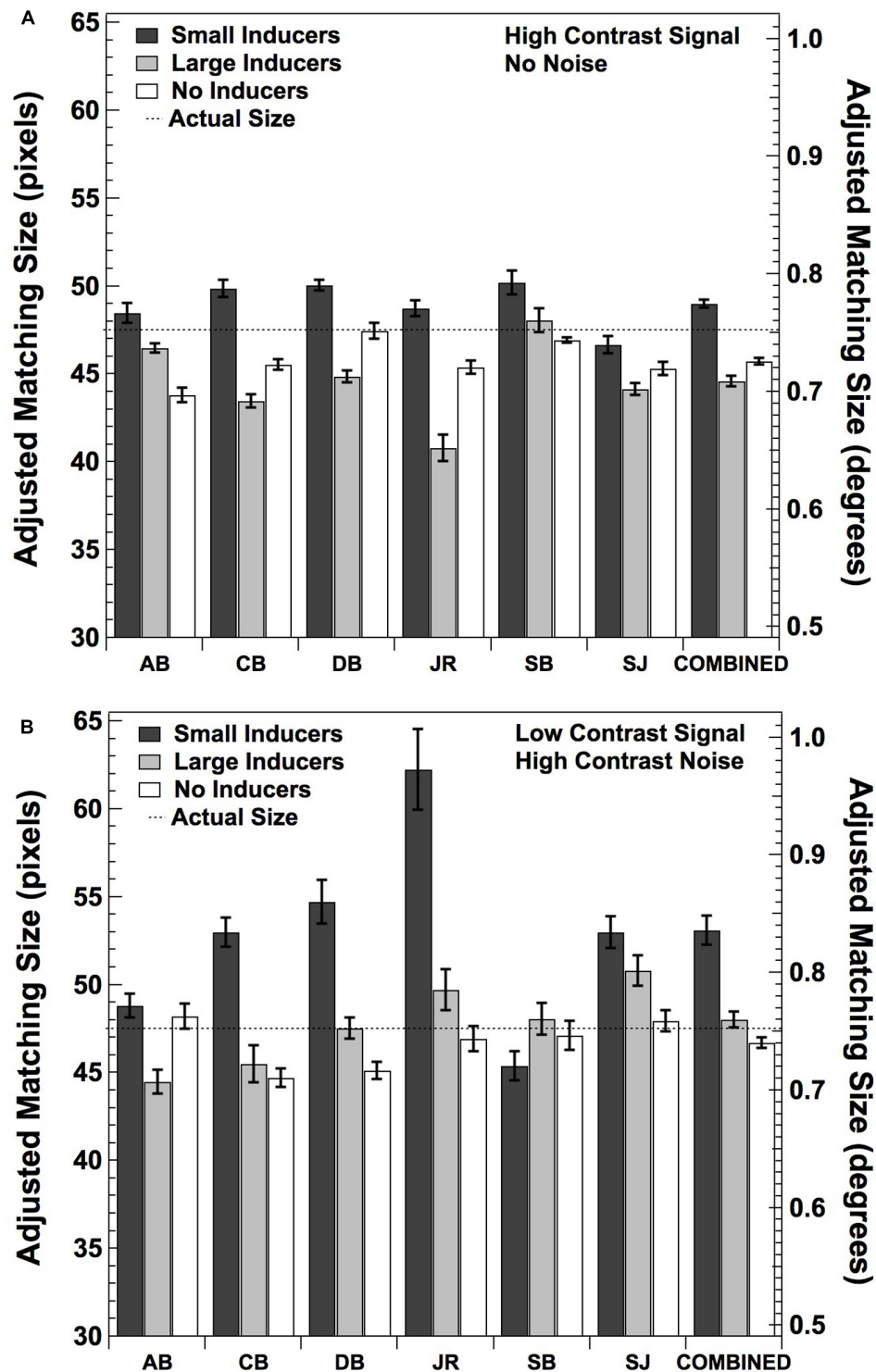
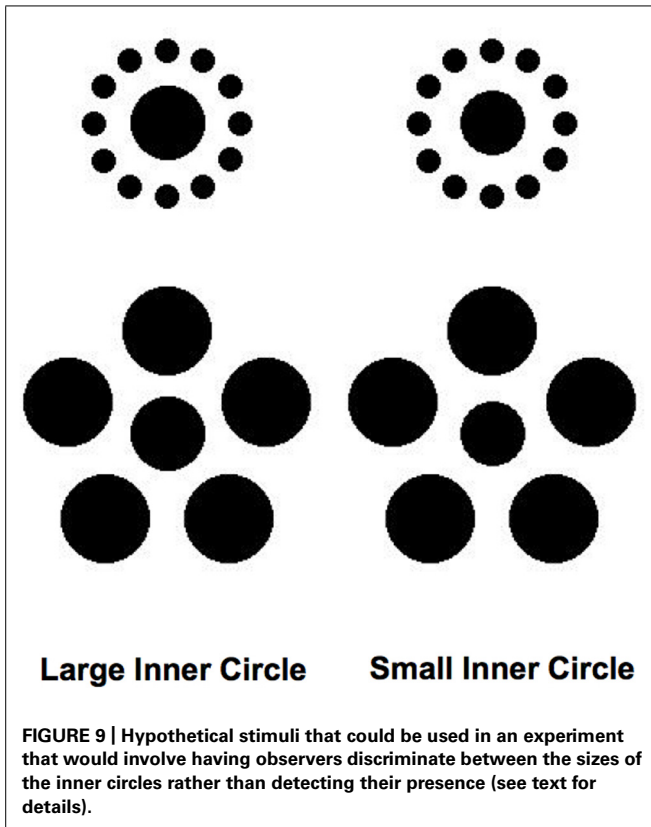


FIGURE 8 | Mean adjusted size matches for six new observers and the data combined across all observers under the same conditions as in Experiment 1 (A) and Experiment 2 (B). Error bars correspond to  $\pm 2$  standard errors of the mean.





observers made use of information within the context of large and small inducers – an effect that is presumably related to the subjective experience of the illusion. However, the results of at least one study suggest that there may in fact be some effect of the relative contrast of the central and surrounding dots in the magnitude of the illusion. Jaeger and Pollack (1977) asked participants to make subjective judgments of the size of the central dot when (a) the inducing dots and the central dot were both “black” and (b) the inducing dots were “black” and the central dot was “gray” (i.e., relatively lower in contrast). Stimuli were shown against a uniform “white” background, and the inducing elements were either larger or smaller than the central dot (the actual luminance or contrast values used in the experiment were not specified). They found that the magnitude of the illusion was reduced when the central dot was gray relative to when it was black when the inducing dots were large; however, they found the opposite effect when the inducing elements were small: the magnitude of the illusion increased when the central dot was gray relative to when it was black.

Although the above study suggests that there may be some relationship between the relative contrasts of the central and surrounding elements and the magnitude of the Ebbinghaus–Titchener illusion, the asymmetric effects of inducer size and the lack of specification of the luminance and contrast levels make the result somewhat difficult to interpret. As such, we decided to address this issue experimentally by having a new set of six observers make subjective size ratings with low contrast stimuli in the presence of high contrast noise, modeled closely after the

conditions experienced by our observers when participating in Experiment 2. Specifically, we averaged the contrast energy thresholds obtained for our original three observers and doubled this value, in order to place it just over detection threshold. We then used this value to set the contrast of the inner dot of the illusion figure, in each of the conditions described in the Experiment 1 (i.e., Large Inducers, Small Inducers, and No Inducers). We also added high contrast Gaussian noise to the figure, in the same manner and at with the same variance as described in Experiment 2. A new sample of noise was added to the figure for every trial of the experiment (15 trials in each condition), and the offset comparison dot that observers were asked to adjust remained high in contrast and noise-free. Each observer was tested in these three conditions, as well as the same three high-contrast, no-noise conditions originally tested in Experiment 1 (six conditions in all). The order of the conditions was randomized for each observer. All other aspects of the experiment were the same as described in Experiment 1.

The results of this subjective rating experiment are shown in Figure 8. Figure 8A shows the results for the conditions that are the same as Experiment 1 (i.e., high contrast stimuli with no added noise). All observers exhibited the characteristic effect of judging the central dot to be relatively greater in size in the context of small than large inducers, and four of the six observers judged the size of the central dot to fall somewhere in between in the absence of inducers. A one-way repeated measures ANOVA revealed a significant effect of condition [ $F(2,5) = 12.53, p < 0.01$ ]. *Post hoc* comparisons using the Tukey HSD test indicated that the mean estimates were significantly greater in the Small Inducers condition than the Large Inducers condition ( $p < 0.01$ ) as well as the No Inducers condition ( $p < 0.05$ ), with no significant difference between the Large Inducers and No Inducers conditions.

Figure 8B shows the results when the middle dot was low in contrast and embedded in high contrast noise. All but one observer (SB) exhibited the characteristic effect of judging the central dot to be relatively greater in size in the context of small than large inducers. Surprisingly, only one observer (AB) judged the size of the central dot to fall somewhere in between these sizes in the absence of inducers; the remaining five observers judged the size of the central dot to be *smallest* in the absence of inducers. This result is consistent with the asymmetric effects of brightness reported by Jaeger and Pollack (1977). A one-way repeated measures ANOVA again revealed a significant effect of condition [ $F(2,5) = 5.51, p < 0.05$ ]. *Post hoc* comparisons using the Tukey HSD test indicated that the mean estimates were significantly greater in the Small Inducers condition than the Large Inducers and No Inducers conditions ( $p < 0.05$ ), with no significant difference between the Larger Inducers and No Inducers conditions. Finally, A 2 (stimulus contrast condition)  $\times$  3 (inducer condition) two-factor ANOVA with repeated measures on both factors showed that there was a significant effect of inducer condition [ $F(5,2) = 11.11, p < 0.01$ ] with no significant effect of stimulus contrast condition [ $F(5,1) = 3.53, p = 0.11$ ] nor a significant inducer condition  $\times$  stimulus contrast condition interaction [ $F(5,2) = 1.09, p = 0.37$ ]. Taken together, these results demonstrate that the Ebbinghaus–Titchener size-contrast illusion

is relatively unaffected by the presentation of the central dot at a low level of contrast within high contrast pixel noise, and strongly suggest that our original three observers were experiencing the size illusion under the conditions used in Experiment 2.

## CONCLUSION

The results of our experiments offer some interesting new insights into the information processing correlates of the Ebbinghaus–Titchener size-contrast illusion. Namely, the subjective size of the central element in the illusion appears to be related to the amount of weight observers assign to the areas within and directly surrounding the central element as well as the range of spatial frequencies that they rely upon when they are asked to perform a simple detection task. We were unable to account for this effect by a simple model that incorporates the overall spatial frequency filtering characteristics of early visual processing, as summarized by the foveal CSF of a normal human adult. Given these results, it may be tempting to conclude that the effects we have observed are due to the operation of processes involved with making higher-level judgments about the relative sizes of objects (e.g., Massaro and Anderson, 1971; Coren and Girgus, 1978; Coren and Enns, 1993). However, it is still possible that a more detailed front-end model (e.g., Chirimuuta et al., 2003) that incorporates additional aspects of the early stages of visual processing, such as oriented V1 receptive fields, parafoveal variations in contrast sensitivity, and cortical magnification, might make predictions not captured by simply incorporating the overall CSF, and these predictions may map more directly on to the results of our classification image analyses.

Finally, although we chose to use a detection task in our experiments for its relative simplicity, an interesting future direction would be to carry out similar experiments using tasks that might rely more directly upon an observer's ability to make judgments about relative size. **Figure 9** illustrates a task and set of stimuli one might use in such a hypothetical experiment. In this case, an observer would be asked to determine which of two central dots that slightly differ in size had appeared on a given trial, in the presence of either large or small inducing elements. It is possible that such a task would tap more directly into the same underlying processes that lead to the misperception of size associated with the subjective experience of the Ebbinghaus–Titchener illusion. We are currently exploring these and other possibilities.

## ACKNOWLEDGMENT

We would like to thank Patrick J. Mundy for his assistance in data collection.

## REFERENCES

- Abbey, C. K., and Eckstein, M. P. (2002). Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J. Vis.* 2, 66–78. doi: 10.1167/2.1.5
- Abbey, C. K., and Eckstein, M. P. (2007). Classification images for simple detection and discrimination tasks in correlated noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 24, B110–B124. doi: 10.1364/JOSAA.24.00B110
- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science* 262, 2042–2044. doi: 10.1126/Science.8266102
- Aglioti, S., DeSouza, J. F., and Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Curr. Biol.* 5, 679–685. doi: 10.1016/S0960-9822(95)00133-3
- Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *J. Vis.* 2, 121–131. doi: 10.1167/2.1.8
- Ahumada, A. J., and Lovell, J. (1971). Stimulus features in signal detection. *J. Acoust. Soc. Am.* 49, 1751–1756. doi: 10.1121/1.1912577
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature* 226, 177–178. doi: 10.1038/226177a0
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897x00357
- Braje, W. L., Tjan, B. S., and Legge, G. E. (1995). Human-efficiency for recognizing and detecting low-pass filtered objects. *Vis. Res.* 35, 2955–2966. doi: 10.1016/0042-6989(95)00071-7
- Busch, A., and Müller, H. J. (2004). The Ebbinghaus illusion modulates visual search for size-defined targets: evidence for preattentive processing of apparent object size. *Percept. Psychophys.* 66, 475–495. doi: 10.3758/BF03194895
- Chirimuuta, M., Clatworthy, P. L., and Tolhurst, D. J. (2003). Coding of the contrasts in natural images by visual cortex (V1) neurons: a Bayesian approach. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1253–1260. doi: 10.1364/JOSAA.20.001253
- Choplin, J. M., and Medin, D. L. (1999). Similarity of the perimeters in the Ebbinghaus illusion. *Percept. Psychophys.* 61, 3–12. doi: 10.3758/BF03211944
- Chung, S. T., Legge, G. E., and Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vis. Res.* 42, 2137–2152. doi: 10.1016/S0042-6989(02)00092-5
- Coren, S., and Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Percept. Psychophys.* 54, 579–588. doi: 10.3758/BF03211782
- Coren, S., and Girgus, J. S. (1978). *Seeing is Deceiving: The Psychology of Visual Illusions*. Hillsdale, NJ: Lawrence Erlbaum Associates (distributed by Halsted Press).
- Cornsweet, T. N. (1970). *Visual Perception*. New York: Academic Press.
- Daneyko, O., Zavagno, D., and Zanuttini, L. (2011). Lightness effects in Delboeuf and Ebbinghaus size-contrast illusions. *Perception* 40, 464–473. doi: 10.1068/p6622
- de Fockert, J., Davidoff, J., Fagot, J., Parron, C., and Goldstein, J. (2007). More accurate size contrast judgments in the Ebbinghaus Illusion by a remote culture. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 738–742. doi: 10.1037/0096-1523.33.3.738
- Eckstein, M. P., Shimozaki, S. S., and Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *J. Vis.* 2, 25–45. doi: 10.1167/2.1.3
- Efron, B., and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall. doi: 10.1007/978-1-4899-4541-9
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314. doi: 10.1037//0033-295x.96.2.267
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X. J., et al. (1999). An anchoring theory of lightness perception. *Psychol. Rev.* 106, 795–834. doi: 10.1037//0033-295x.106.4.795
- Girgus, J. S., Coren, S., and Agdern, M. (1972). The interrelationship between the Ebbinghaus and Delboeuf illusions. *J. Exp. Psychol.* 95, 453–455. doi: 10.1037/h0033606
- Goodale, M. A., and Humphrey, G. K. (1998). The objects of action and perception. *Cognition* 67, 181–207. doi: 10.1016/S0010-0277(98)00017-1
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Haffenden, A. M., Schiff, K. C., and Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: nonillusory effects of pictorial cues on grasp. *Curr. Biol.* 11, 177–181. doi: 10.1016/S0960-9822(01)0023-9
- Handlovsky, I., Hansen, S., Lee, T. D., and Elliott, D. (2004). The Ebbinghaus illusion affects on-line movement control. *Neurosci. Lett.* 366, 308–311. doi: 10.1016/j.neulet.2004.05.056
- Im, H. Y., and Chong, S. C. (2009). Computation of mean size is based on perceived size. *Atten. Percept. Psychophys.* 71, 375–384. doi: 10.3758/APP.71.2.375
- Jaeger, T. (1978). Ebbinghaus illusions: size contrast or contour interaction phenomena? *Percept. Psychophys.* 24, 337–342. doi: 10.3758/BF03204250

- Jaeger, T., and Pollack, R. H. (1977). Effect of contrast level and temporal order on the Ebbinghaus circles illusion. *Percept. Psychophys.* 21, 83–87. doi: 10.3758/BF03199473
- Jameson, D., and Hurvich, L. M. (1964). Theory of brightness and color contrast in human vision. *Vis. Res.* 4, 135–154. doi: 10.1016/0042-6989(64)90037-9
- Johansson, G. (1973). Visual-perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/BF03212378
- Johnston, J. C., and McClelland, J. I. (1974). Perception of letters in words – seek not and ye shall find. *Science* 184, 1192–1194. doi: 10.1126/Science.184.4142.1192
- Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. New York: Praeger.
- Koffka, K. (1935). *Principles of Gestalt Psychology*. New York: Harcourt, Brace and Company.
- Lotto, R. B., and Purves, D. (2000). An empirical explanation of color contrast. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12834–12839. doi: 10.1073/pnas.210369597
- Massaro, D. W., and Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. *J. Exp. Psychol.* 89, 147–151. doi: 10.1037/h0031158
- Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
- Muller, H. J., and Busch, A. (2006). Visual search for size-defined target objects is modulated by the Ebbinghaus apparent-size illusion: facilitatory and inhibitory effects of the context objects. *Perception* 35, 671–700. doi: 10.1068/p5139
- Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2002). Optimal methods for calculating classification images: weighted sums. *J. Vis.* 2, 79–104. doi: 10.1167/2.1.6
- Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2005). Classification images predict absolute efficiency. *J. Vis.* 5, 139–149. doi: 10.1167/5.2.5
- Nandy, A. S., and Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *J. Vis.* 8:3.1–3.20. doi: 10.1167/8.13.3
- Pavlova, M., and Sokolov, A. (2000). Speed perception is affected by the Ebbinghaus–Titchener illusion. *Perception* 29, 1203–1208. doi: 10.1068/p3047
- Phillips, W. A., Chapman, K. L., and Berry, P. D. (2004). Size perception is less context-sensitive in males. *Perception* 33, 79–86. doi: 10.1068/p5110
- Pomerantz, J. R., Sager, L. C., and Stoever, R. J. (1977). Perception of wholes and of their component parts: some configural superiority effects. *J. Exp. Psychol. Hum. Percept. Perform.* 3, 422–435. doi: 10.1037/0096-1523.3.3.422
- Roberts, B., Harris, M. G., and Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception* 34, 847–856. doi: 10.1068/p5273
- Rose, D., and Bressan, P. (2002). Going round in circles: shape effects in the Ebbinghaus illusion. *Spat. Vis.* 15, 191–203. doi: 10.1163/15685680252875165
- Shimozaki, S. S., Eckstein, M. P., and Abbey, C. K. (2005). Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. *J. Vis.* 5, 45–57. doi: 10.1167/5.1.5
- Sperandio, I., Lak, A., and Goodale, M. A. (2012). Afterimage size is modulated by size-contrast illusions. *J. Vis.* 12:18.1–18.10. doi: 10.1167/12.2.18
- Sperandio, I., Savazzi, S., and Marzi, C. A. (2010). Is simple reaction time affected by visual illusions? *Exp. Brain Res.* 201, 345–350. doi: 10.1007/s00221-009-2023-y
- Teller, D. Y. (1984). Linking propositions. *Vis. Res.* 24, 1233–1246. doi: 10.1016/0042-6989(84)90178-0
- Titchener, E. B. (1901). *Experimental Psychology: A Manual of Laboratory Practice*. New York: The Macmillan Company.
- Tjan, B. S., Braje, W. L., Legge, G. E., and Kersten, D. (1995). Human-efficiency for recognizing 3-D objects in luminance noise. *Vision Res.* 35, 3053–3069. doi: 10.1016/0042-6989(95)00070-G
- Wallach, H., and O'Connell, D. N. (1953). The kinetic depth effect. *J. Exp. Psychol.* 45, 205–217. doi: 10.1037/h0056880
- Watson, A. B. (2000). Visual detection of spatial contrast patterns: evaluation of five simple models. *Opt. Express* 6, 12–33. doi: 10.1364/OE.6.000012
- Weintraub, D. J. (1979). Ebbinghaus illusion: context, contour, and age influence the judged size of a circle amidst circles. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 353–364. doi: 10.1037/0096-1523.5.2.353
- Weintraub, D. J., and Schneck, M. K. (1986). Fragments of Delboeuf and Ebbinghaus illusions: contour/context explorations of misjudged circle size. *Percept. Psychophys.* 40, 147–158. doi: 10.3758/BF03203010
- Westwood, D. A., and Goodale, M. A. (2003). A haptic size-contrast illusion affects size perception but not grasping. *Exp. Brain Res.* 153, 253–259. doi: 10.1007/s00221-003-1599-x

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 November 2013; paper pending published: 06 December 2013; accepted: 04 February 2014; published online: 19 February 2014.

Citation: Gold JM (2014) Information processing correlates of a size-contrast illusion. *Front. Psychol.* 5:142. doi: 10.3389/fpsyg.2014.00142

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Gold. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Phase noise reveals early category-specific modulation of the event-related potentials

Kornél Németh<sup>1</sup>, Petra Kovács<sup>1</sup>, Pál Vakli<sup>1</sup>, Gyula Kovács<sup>1,2,3</sup> and Márta Zimmer<sup>1\*</sup>

<sup>1</sup> Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary

<sup>2</sup> DFG Research Unit Person Perception, Friedrich Schiller University of Jena, Jena, Germany

<sup>3</sup> Institute of Psychology, Friedrich Schiller University of Jena, Jena, Germany

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Frédéric Gosselin, University of Montreal, Canada

Daniel Hart Baker, University of York, UK

## \*Correspondence:

Márta Zimmer, Department of Cognitive Science, Budapest University of Technology and Economics, Eötvös József utca 1., Budapest 1111, Hungary  
e-mail: mzimmer@cogsci.bme.hu

Previous studies have found that the amplitude of the early event-related potential (ERP) components evoked by faces, such as N170 and P2, changes systematically as a function of noise added to the stimuli. This change has been linked to an increased perceptual processing demand and to enhanced difficulty in perceptual decision making about faces. However, to date it has not yet been tested whether noise manipulation affects the neural correlates of decisions about face and non-face stimuli similarly. To this end, we measured the ERPs for faces and cars at three different phase noise levels. Subjects performed the same two-alternative age-discrimination task on stimuli chosen from young–old morphing continua that were created from faces as well as cars and were calibrated to lead to similar performances at each noise-level. Adding phase noise to the stimuli reduced performance and enhanced response latency for the two categories to the same extent. Parallel to that, phase noise reduced the amplitude and prolonged the latency of the face-specific N170 component. The amplitude of the P1 showed category-specific noise dependence: it was enhanced over the right hemisphere for cars and over the left hemisphere for faces as a result of adding phase noise to the stimuli, but remained stable across noise levels for cars over the left and for faces over the right hemisphere. Moreover, noise modulation altered the category-selectivity of the N170, while the P2 ERP component, typically associated with task decision difficulty, was larger for the more noisy stimuli regardless of stimulus category. Our results suggest that the category-specificity of noise-induced modulations of ERP responses starts at around 100 ms post-stimulus.

**Keywords:** phase noise, category effect, P1, N170, P2

## INTRODUCTION

There has been a long tradition of applying external noise to visual stimuli in the last two decades of the 20th century in visual psychophysics as well as in studies of face perception to study various stages of visual processing (Costen et al., 1994; Gold et al., 1999; Näsänen, 1999). Common methods included noise manipulation combined with electrophysiological and brain imaging methods to investigate and identify the underlying neuronal mechanisms of the various functions of the perceptual system. In recent studies, different types of external noise were used, including uniform white noise (Wild and Busey, 2004), Gaussian noise (Jemel et al., 2003), bit noise (Smith et al., 2012), multiplicative noise combined with brain imaging techniques (e.g., Schyns et al., 2003, 2007, 2009; Smith et al., 2004, 2006, 2007, 2008, 2009; Rutishauser et al., 2011), Fourier phase-randomization techniques (Rousselet et al., 2008a; Bankó et al., 2011) with the mean-phase randomization (Dakin et al., 2002), and pink noise (Tjan et al., 2006; Rousselet et al., 2008a,b). These techniques provided valuable insights into the spatial and temporal events at different cortical regions in the human brain involved in different stages of face processing.

Regarding human face perception, electrophysiological studies have described a large positive (P1) and negative (N170) wave

over the occipital and posterior occipito-temporal areas that might be sensitive to face stimulation (Bentin et al., 1996; Eimer, 2000a; Itier and Taylor, 2004). As of today, usually the N170 is considered as the first clearly face-sensitive event-related potential (ERP) component, although category-specific processes have been suggested by some studies to be present already at 100 ms (or even 50–80 ms) after stimulus onset (corresponding to the P1 component; George et al., 1997; Seeck et al., 1997; Liu et al., 2002; Herrmann et al., 2005a; Thierry et al., 2007). The N170 is higher in amplitude and shorter in latency to pictures of faces than to exemplars of other non-face object categories (Bentin et al., 1996; for reviews see Rossion and Jacques, 2008, 2011; Eimer, 2011). Recently, however, the specificity of N170 for faces has been questioned by studies that failed to demonstrate higher N170 amplitude for faces when compared with cars (Rossion et al., 2000a; Schweinberger et al., 2004; Thierry et al., 2007; Dering et al., 2011; Kloth et al., 2013).

With regard to noisy stimulation, Jemel et al. (2003) used a parametric design to characterize early ERPs to face stimuli embedded in gradually decreasing levels of random Gaussian noise. The authors found that while the P1 component was unaffected by noise levels, there was a linear increase in the amplitude and a decrease in the latency of the N170 with decreasing levels of



noise. Jemel et al. (2003) concluded that while the early P1 component is likely to reflect the stage at which the perceptual analysis of faces is achieved, the N170 seems to reflect the successful categorization of faces (Liu et al., 2002; Jemel et al., 2003). In other words, earlier ERP components might reflect the extraction of task-relevant information from noisy stimuli. This modulation of the N170 component is in line with findings showing attenuated and delayed N170 to faces either without internal features or in the absence of their contours (Eimer, 2000a). In addition, Rousselet et al. (2008b) found that sensitivity to phase noise falls in the time window of the N170 (130–170 ms).

The P2 ERP component is characterized by a positive-going deflection over lateral occipito-temporal areas and a maximal peak between 200 and 250 ms. Recently, it has been shown that the amplitude of the P2 is sensitive to the inversion of either the entire face or of its parts (Milivojevic et al., 2003; Boutsen et al., 2006) and has been linked to the processing of spatial relations between facial features in individual faces (Latinus and Taylor, 2006). Wiese et al. (2009) have shown that own race faces generate larger P2 components when compared with faces of other races, although, this effect interacts with expertise (Stahl et al., 2008). Larger P2 was also reported for younger when compared to older face stimuli (Stahl et al., 2008). Furthermore, it has been suggested that the P2 is involved in individual face recognition mechanisms (Halit et al., 2000). Altogether, these results suggest that the P2 is involved in the deeper and more advanced analysis of faces when compared to earlier components. Regarding noisy stimulation, Rousselet et al. (2007, 2008a) showed that the P2 is larger to noise patterns in comparison to faces. In a follow-up study, they tested whether this difference was independent from the changes of the N170 amplitude and therefore a peak-to-peak analysis was carried out on the modeled data (Rousselet et al., 2008b). The authors found that the P2 difference is a simple carry-over effect that was present already on the N170. In addition, the P2 was identified as a clear neural correlate of decision difficulty under noisy stimulation (Philiastides et al., 2006; Heekeren et al., 2008). However, a recent study using image warping as well as phase noise to manipulate task difficulty found that rather, the P2 reflects noise-sensitive increases of sensory processing and not task difficulty *per se* (Bankó et al., 2011). In a previous ERP study, we confirmed these results and distinguished the nature of adding phase noise from that of another irrelevant, overlapping car image (Nagy et al., 2009). We found that adding phase noise reduces the N170 component, while the amplitude of the P2 component increases with the amount of noise added. In addition, the P2 was larger in the phase noise condition than in another coherent, but irrelevant stimulus (a car) was added to the face.

In general, adding noise to face images leads to smaller N170 amplitudes, reflecting impaired early structural face processing (Bentin and Deouell, 2000; Eimer, 2000a,b for a review see Rossion and Jacques, 2008), as well as to larger P2 amplitudes. However, the effect of noise reflected in the early P1 component is equivocal as of today. While Jemel et al. (2003) found that the effect of added noise does not affect P1 amplitude, other studies have demonstrated that the P1 and P2 components are significantly larger in the noise-present when compared with noise-absent conditions

(e.g., Curran et al., 1993; Tucker et al., 1994; Mercure et al., 2008; Bankó et al., 2011).

To the best of our knowledge, so far no study has explicitly compared the noise-dependence of face and non-face stimulus categories. The goal of the present study was to test whether adding phase noise to stimuli affects the neural processing of different high-level categories, such as faces and cars, in a similar way.

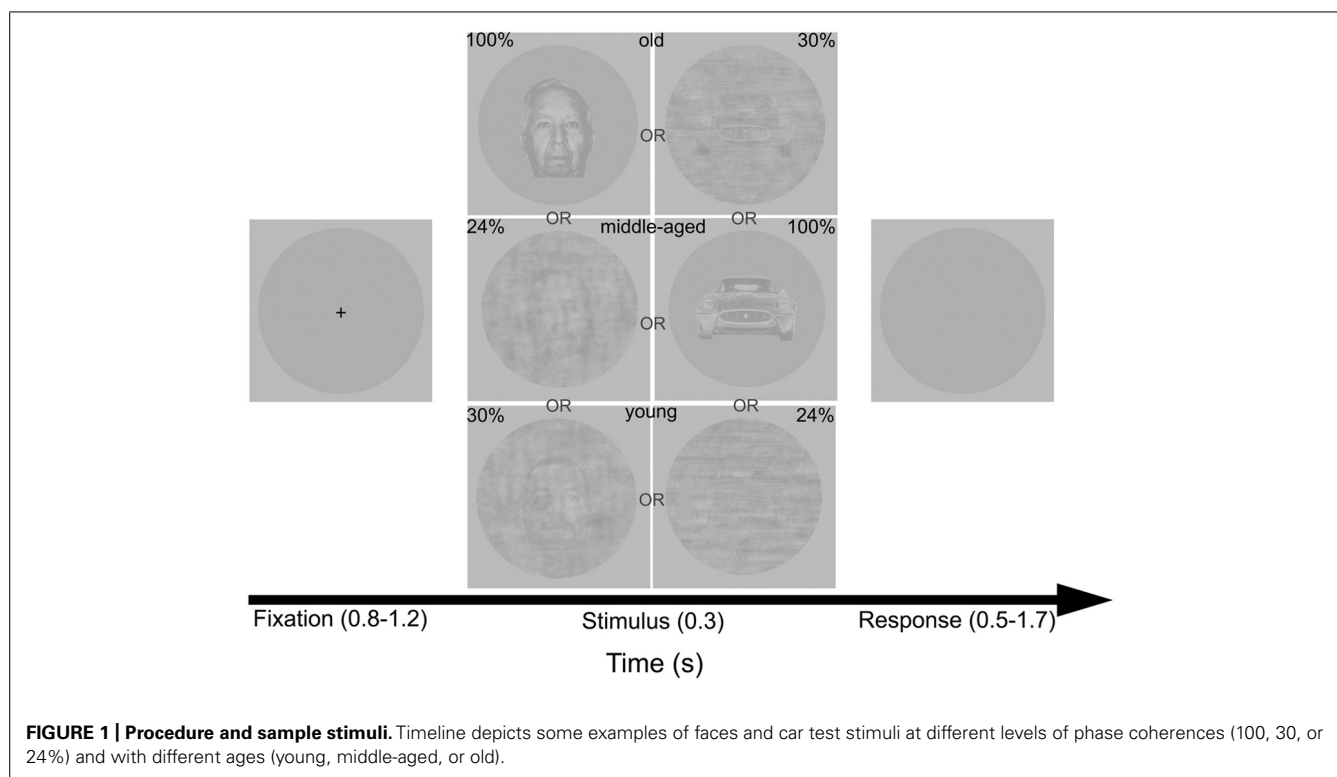
## MATERIALS AND METHODS

### PARTICIPANTS

Sixteen naïve, healthy volunteers (two left-handed, eight females, mean age: 22.1 years  $\pm$  2.1 years SD) participated in the study. They received partial course credits for their participation and gave signed, informed consent in accordance with the Ethical Committee of the Budapest University of Technology and Economics prior to testing. All participants had normal or corrected-to-normal visual acuity, no previous history of any neurological or ophthalmologic diseases and were not under medication. Three participants were excluded from the final electrophysiological analyses due to insufficient numbers of ERP segments after artifact rejection. Therefore, statistical analysis was conducted on the data of thirteen subjects (seven females, one left-handed, mean age: 21.5 years  $\pm$  1.8 years SD).

### STIMULI

Front-view grayscale images of faces and cars were used with age gradually changing, with or without phase noise. Face stimuli were digital images of six Caucasian males from a larger face database (Minear and Park, 2004). Three of them were younger than 30 years old, while the others were older than 60 years old. Car images were old and new variations of the same models of three well-known commercial car types (VW, Mercedes, and Jaguar), and were downloaded from freely available websites. Car images were presented in full frontal views, similar to those of Kloth et al. (2013). All images were first converted into grayscale (8 bit) using Adobe Photoshop CS3 Extended 10.0 (Adobe Systems Inc.). Stimuli of both categories were then revealed through a circular aperture (radius = 153 pixels). Stimulus size was equated for each category (mean height and width of the faces and cars were 248 and 154 pixels, and 153 and 251 pixels, respectively; see **Figure 1**). Since previous studies have shown that early ERP components, such as P1, are sensitive to luminance (Johannes et al., 1995) and that neural processes are sensitive to luminance histogram skewness (Olman et al., 2008), we have equated all stimuli in luminance and matched their histograms using the *lummatch* and *histmatch* functions of the SHINE toolbox (Willenbockel et al., 2010). On the other hand, we did not equate the spectral content of the images, as we would concurrently have manipulated artificially the difficulty of the age-discrimination task for the face stimuli. It is well known that facial aging is reflected in the dynamic, cumulative effects of the skin, and is a complex synergy of skin textural changes and the loss of facial volume (Coleman and Grover, 2006). The decreased tissue elasticity and the redistribution of subcutaneous fullness result in a larger amount of higher spatial frequency information. This low-level difference between younger/newer and older individuals does not appear when comparing new cars to old ones.



In order to increase task difficulty, two different types of stimulus manipulations were applied. First, we decreased the age difference between young and old stimuli using a warping algorithm (Winmorph 3.01; Kovács et al., 2005, 2006, 2007; Bankó et al., 2011). That is, we paired a young and an old image of the same category and created a morph continuum with seven intermediate images of faces and cars. Second, the coherence of the original images (100% phase coherence) and the intermediate morphs was manipulated by decreasing their phase coherence in two steps (30 and 24% phase coherences, respectively) using the weighted mean phase technique (Dakin et al., 2002). In fact it means that we have manipulated the phase coherence of the RGB values (and not the luminance values) of the stimuli. This phase-randomization resulted in the gradual elimination of the cues important for accurate age judgments.

To avoid behavioral ceiling or floor-effects and to have comparable performance for face and car stimuli, first we performed a behavioral pilot experiment ( $n = 12$ ). We tested the age discrimination performance of participants for 10 exemplars of faces and cars as well as for 10 incrementally graded noise levels from 0 to 100% phase coherence. For the final three stimulus-pairs, morph levels of the young-old continuum and the exact percentage of phase noise were selected based on the results of this pilot study, so that the average age-discrimination performance would be similar across faces and cars for each phase noise level.

Stimuli were presented centrally on a uniform gray background on a 26 inch LCD monitor at a refresh rate of 60 Hz, while viewing distance (57 cm) was maintained using a chinrest. Stimulus

presentation was controlled by MATLAB 2008a (Mathworks, Natick, MA, USA) using Psychtoolbox 3.0.9 (Brainard, 1997; Pelli, 1997) and custom-made scripts.

## PROCEDURE

As it is generally more difficult to determine the age of a car than the age of a face, as suggested by the results of the pilot study, first, participants were presented with a practice session for the car stimuli prior to the experiment.

### Practice experiment

In the first part of the practice, participants had to choose the younger (newer) car from a pair of stimuli, depicting the end-points of the morph continuum, or in other words the oldest and youngest versions of a model. Each pair was presented eight times (exposition time = until response; inter-trial interval = 500 ms). The newer model was displayed randomly on either the left or the right side. Participants received feedback after each trial as well as at the end of the block. Participants performed at least four, but not more than six blocks of 24 trials. The practice was interrupted if 90% correct performance was reached in two consecutive blocks. A subject was excluded from the study if their performance did not reach this criterion even after 10 practice blocks (0 participants).

Second, participants performed an age-discrimination task on individually presented cars depicting the endpoints of morph continuums. In this part of the practice a fixation screen was presented in the beginning of each trial for a random time between 800 and 1200 ms, followed by the presentation of the test image (100% phase coherence) for 300 ms. Participants were

instructed to respond within 2 s after stimulus onset (inter-trial interval = 800 ms). Within a single block, car stimuli were presented in a random order. Subjects had to perform 4–6 blocks of 24 trials (each car presented four times in a random order). The practice was interrupted if 90% correct performance was reached in two consecutive blocks. A subject was excluded from the study if her/his performance did not reach this criterion even after six practice blocks (0 participants).

Finally, immediately prior to the ERP recording experiment, participants were asked to passively fixate the center of each stimulus (both faces and cars) at each noise level and at each morph level for 5000 ms (inter-stimulus interval 1000 ms) for the subjects once, to avoid strong familiarity effects of the practice phase with cars.

### ERP recording experiment

Subjects performed an old vs. young age discrimination task for faces and cars. The trial structure was identical to that of the second task of the practice experiment (**Figure 1**). Noise-levels, stimulus categories, and morph levels were intermixed and presented in random order within each block. Each participant completed eight blocks of 378 trials [ $2(\text{category}; \text{face vs. car}) \times 3(\text{exemplars}; \text{face morph-continuum vs. car morph-continuum}) \times 3(\text{coherence level}, 100\% \text{ vs. } 30\% \text{ vs. } 24\%) \times 7(\text{morph level}) \times 3(\text{number of repetitions})$ ]. Subjects were allowed to take a short break between blocks. An experimental session lasted approximately 100 min.

### BEHAVIORAL DATA ANALYSIS

Accuracy and response times (RTs) were collected during the experiment. Performance was assessed by computing just noticeable differences (JND) as the smallest difference in morph level required to perform the old versus young age discrimination task reliably (Lee and Harris, 1996; Bankó et al., 2009) for each stimulus type individually. First, psychophysical data were modeled by the cumulative Gaussian psychometric function, using the *Psignifit* toolbox (Version 2.5.6.) for MATLAB (Wichmann and Hill, 2001). JNDs were calculated using the equation  $\text{JND} = (\text{Perf}_{75} - \text{Perf}_{25})/2$ , where  $\text{Perf}_{75}$  and  $\text{Perf}_{25}$  denote the morph levels leading to 75 and 25% accuracies, respectively. JNDs for different stimuli and noise levels were calculated separately. RTs were calculated as the average of the RTs for stimuli yielding 25 and 75% performance. JNDs and RTs were analyzed with a  $2 \times 3$  repeated measures ANOVA with category (2; face vs. car) and phase coherence (100% vs. 30% vs. 24%) as within-subject factors. *Post hoc t*-tests were computed using Fisher's Least Significant Difference (LSD) tests.

### ELECTROPHYSIOLOGICAL RECORDING AND ANALYSIS

#### EEG acquisition and processing

Electroencephalography (EEG) data was recorded using a BrainAmp (BrainProducts GmbH, Munich, Germany) amplifier from 60 Ag/AgCl scalp electrodes placed according to the international 10/10 electrode system (Chatrian et al., 1985) and mounted on an ActiCap (Easycap, Herrsching-Breitbrunn, Germany). Additionally, four periocular electrodes were placed at the outer canthi of the eyes and above and below the right eye for recording the

electrooculogram (EOG). All channels were referenced to FCz online and digitally transformed to a common averaged reference offline. The ground was placed at AFz and all input impedances were kept below 10 k $\Omega$ . EEG was digitized at a 1000 Hz sampling rate with an analog bandpass filter of 0.016–1000 Hz. Subsequently, a digital 0.1 Hz, 12 dB/octave Butterworth zero phase high-pass filter was used to remove DC shifts, and a 50 Hz notch filter was applied to minimize line-noise artifacts. Finally, a 12 dB/octave low-pass filter with a cut-off frequency of 50 Hz was applied. Trials that contained voltage fluctuations exceeding  $\pm 100 \mu\text{V}$ , or eye blinks exceeding  $\pm 50 \mu\text{V}$  were rejected.

#### ERP data analysis

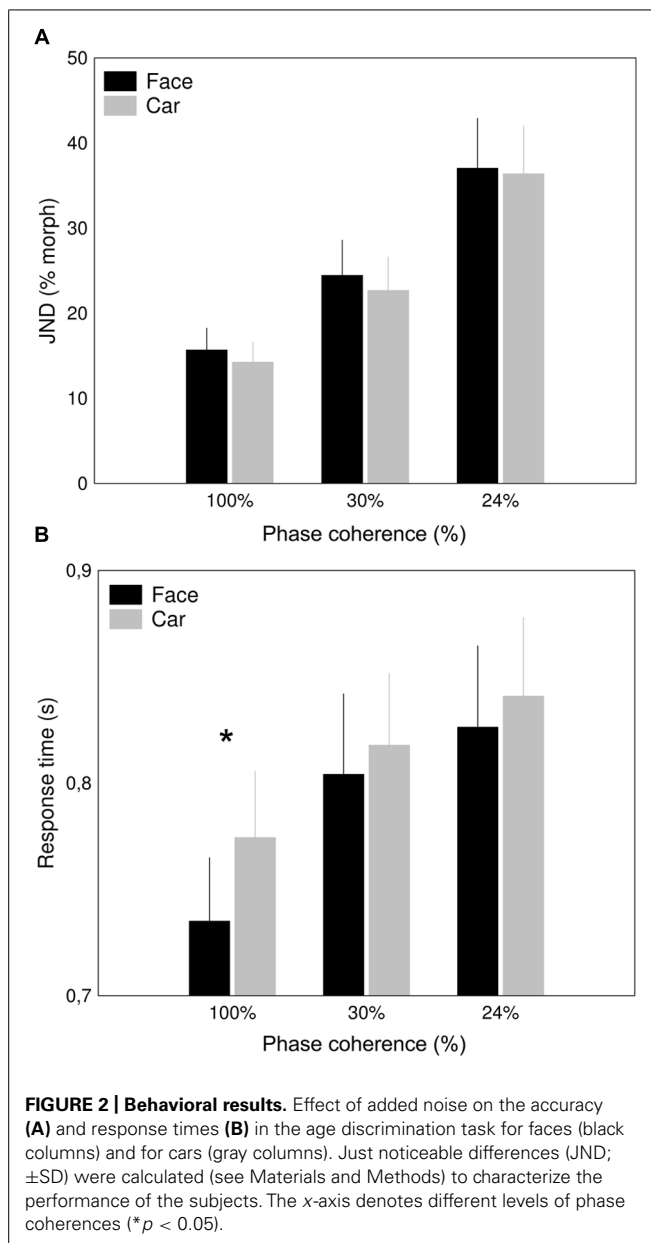
After the eye blink artifacts were corrected (Gratton et al., 1983) the EEG was segmented offline using Brain Vision Analyzer 1.05.0002 (Brain Products GmbH, Munich, Germany) into 1300 ms epochs using a 500 ms pre stimulus interval. Segments were baseline corrected over the 500 ms prestimulus window, artifact rejected, and averaged to obtain the ERP waveforms for each subject and for each condition. Individual ERPs were averaged to compute the grand average ERP for visualization. Statistical analysis was performed on the early visual components P1, N170, and P2 of the individual average ERP waveform. The peak amplitude and latency of the individually averaged ERPs was extracted using a semiautomatic detection algorithm that identified the global maxima separately for each selected channels in a specific time window. P1 was defined as a main positive deflection in the 80–130 ms time window. N170 was defined as a negative component at around 130–200 ms after stimulus onset, and P2 as a second positive component in the 200–250 time window. P1 amplitude was measured over O1, PO7 (left hemisphere, LH), and O2, PO8 (right hemisphere, RH) electrode positions. In the case of the N170, the usual posterior-occipito-temporal sites, corresponding to the PO7, PO9, P7, and P9 (LH) and PO8, PO10, P8, and P10 (RH) were used, while P2 amplitude was measured over PO3, PO7, O1 (LH), and PO4, PO8, and O2 (RH) channels. Both amplitude and latency values of the pooled values of the relevant electrodes were entered into a four-way repeated-measures ANOVA with hemisphere (2; left vs. right), category (2; face vs. car), coherence (3; 100% vs. 30% vs. 24% phase coherence), and age (3; young/new vs. middle-aged vs. old) as within-subject factors separately for each component. The Greenhouse–Geisser correction was applied to correct for possible violations of sphericity. *Post hoc* tests were computed using Fisher's LSD tests.

### RESULTS

#### BEHAVIORAL RESULTS

The age-discrimination performance of the participants was similar for faces and cars (main effect of category:  $F(1,15) = 0.198$ ,  $p = 0.661$ ,  $\eta^2 = 0.013$ ; **Figure 2A**), suggesting that the difficulty of the task was similar for the two stimulus categories. As expected, additional phase noise reduced the performance incrementally (main effect of coherence:  $F(1.11,16.58) = 13.002$ ,  $p < 0.0001$ ,  $\eta^2 = 0.464$ ). This effect was similar for the two stimulus categories, as suggested by the lack of interaction between category and coherence ( $F(2,30) = 0.0461$ ,  $p = 0.955$ ,  $\eta^2 = 0.003$ ).





Paralleling performance results, RTs were also prolonged by reduced phase-coherence (main effect of coherence:  $F(1.1, 16.54) = 23.98$ ,  $p < 0.0001$ ,  $\eta^2 = 0.62$ , **Figure 2B**). In addition, significantly longer RTs were found for car stimuli when compared to faces [main effect of category:  $F(1, 15) = 5.47$ ,  $p = 0.03$ ,  $\eta^2 = 0.27$ ], at least for the 100% coherence condition (category  $\times$  coherence level interaction: [ $F(2, 30) = 3.316$ ,  $p < 0.05$ ,  $\eta^2 = 0.181$ ]).

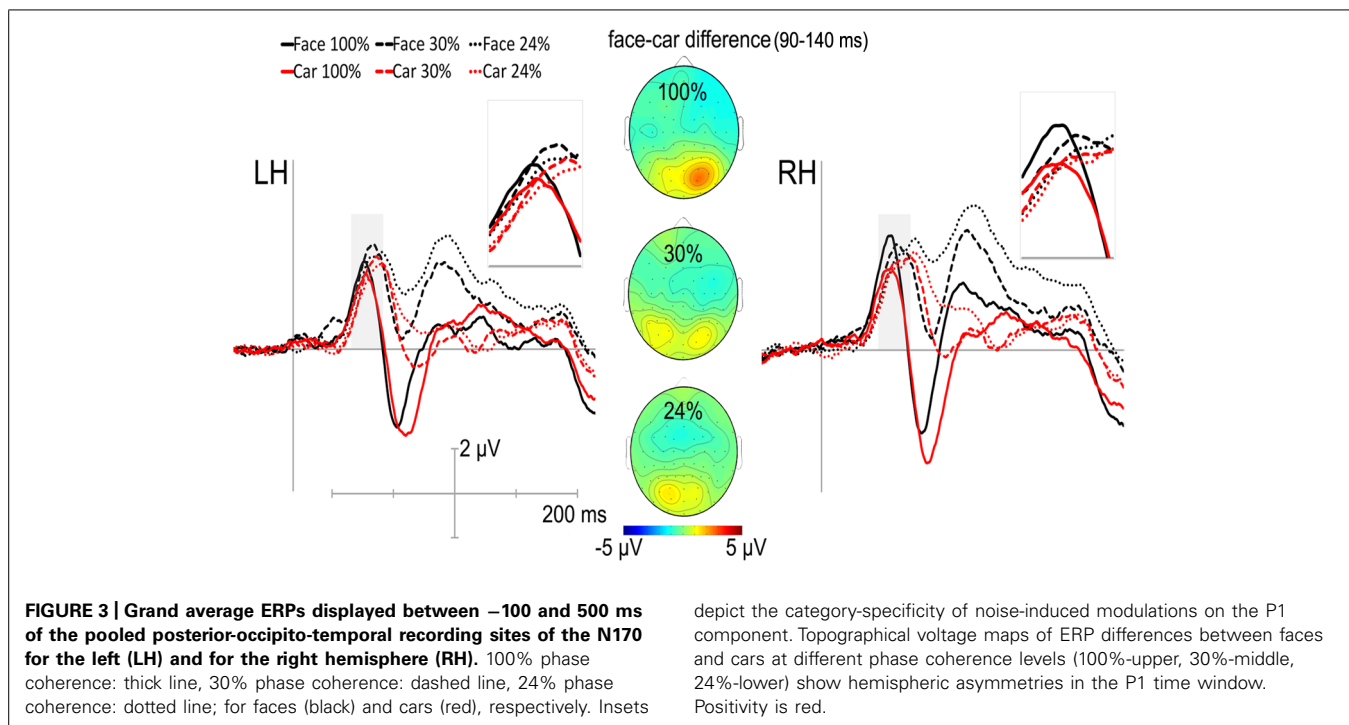
## RESULTS OF THE ELECTROPHYSIOLOGICAL MEASUREMENT

The stimuli evoked ERPs with clearly identifiable P1, N170, and P2 components, measured at occipital and posterior-occipito-temporal sites. **Figure 3** depicts the grand average ERPs of the pooled recording sites over the LH and RH, displayed between  $-100$  and  $500$  ms.

## P1

Significantly larger P1 amplitudes were observed for faces when compared to car stimuli [main effect of category:  $F(1, 12) = 10.16$ ,  $p = 0.008$ ,  $\eta^2 = 0.46$ ]. Importantly, the noise-induced modulation of the P1 component showed category-specificity in a hemisphere-specific manner [hemisphere  $\times$  category  $\times$  coherence interaction:  $F(2, 24) = 8.8452$ ,  $p < 0.01$ ,  $\eta^2 = 0.4243$ ], as it was enhanced as a result of adding noise to the images over the RH for cars (*post hoc* test for 100% vs. 30 or 24%:  $p < 0.005$  for both comparisons) and over the left hemisphere for faces (*post hoc* tests for 100% vs. 30 or 24%:  $p < 0.01$  for both comparisons) but remained stable across phase coherence levels for cars over the left and for faces over the right hemisphere (**Figure 4A**). Facial aging is mainly reflected in changes of skin textures and in altered tissue elasticity. As these changes can increase the amount of higher spatial frequency information only in the case of older face stimuli such low-level differences might explain the different phase-noise dependency of P1 for faces and cars. However, since age decisions for faces are mainly based on these factors (e.g., George and Hole, 2000), we have not equated the spectral content of the images. However, to test whether the significant hemisphere  $\times$  category  $\times$  coherence interaction is due to any differences in the spatial frequency content in the 100% phase coherent stimuli, we tested the effect of wrinkling/skin texture changes on the range of higher spatial frequency information. We plotted the spectral content of the 100% phase coherent stimuli by using the *splot* method of the SHINE toolbox (Willenbockel et al., 2010) and compared these functions for faces and cars at every morph level. Due to the small sample size, we used non-parametric ranked *t*-tests (point-by-point two-tailed Mann–Whitney *U* tests with Bonferroni-corrected *p* values). Although we found that the older the face stimuli, the more pronounced the spectral difference in the range of higher spatial frequency information when compared with car stimuli, it is worth noting that the spectral content of the youngest stimuli did not differ between the two categories. Next, we investigated the hemisphere  $\times$  category  $\times$  coherence  $\times$  age interaction. The results suggest that the age information of the stimuli do not modulate the strength of the hemisphere-specific category effect reflected in the P1 component [hemisphere  $\times$  category  $\times$  coherence  $\times$  age interaction:  $F(4, 48) = 0.33$ ,  $p = 0.86$ ,  $\eta^2 = 0.03$ , n.s.], arguing against the role of low-level spectral differences in explaining the results. Moreover, the age of the stimuli as a categorical factor neither had a main effect [ $F(2, 24) = 2.78$ ,  $p = 0.09$ ,  $\eta^2 = 0.19$ ] nor had any significant two-way (any  $ps > 0.13$ ), three-way (any  $ps > 0.25$ ), or four-way interactions (any  $ps > 0.75$ ) with other factors. Taken together with the fact that no significant differences in spectral content were observed between the youngest 100% phase coherent face and car stimuli, our results suggest that the observed hemisphere  $\times$  category  $\times$  coherence three-way interaction is not due to the low-level spectral differences in the original stimuli.

The latency of the P1 was significantly longer for cars when compared to faces [main effect of category:  $F(1, 12) = 22.65$ ,  $p = 0.0005$ ,  $\eta^2 = 0.65$ ]. Adding phase noise to the stimuli increased the latencies of P1 component [main effect of coherence:  $F(1.27, 15.27) = 9.7$ ,  $p = 0.0008$ ,  $\eta^2 = 0.4468$ , *post hoc* LSD: 100% vs. 30 and 24%:  $p < 0.002$  for both comparison].



This difference in latency was, however, similar for both categories [category  $\times$  coherence interaction:  $F(1.71, 20.5) = 1.77$ ,  $p = 0.19$ ,  $\eta^2 = 0.13$ ; **Figure 4D**].

### N170

We found a significant main effect of coherence for the amplitude of the N170 component, [ $F(1.48, 17.78) = 71.45$ ,  $p < 0.0001$ ,  $\eta^2 = 0.86$ ] reflecting the reduction of the N170 amplitude as the phase coherence decreases. It is worth noting that this effect was larger for the right when compared with the left hemisphere as suggested by the significant hemisphere  $\times$  coherence interaction [ $F(1.28, 15.38) = 5.52$ ,  $p = 0.01$ ,  $\eta^2 = 0.32$ , **Figure 4B**]. Interestingly, N170 amplitudes did not show the typically observed face-specificity (see Kloth et al., 2013 for similar results): the N170 was almost identical for both faces and cars [main effect of category:  $F(1, 12) = 0.49$ ,  $p = 0.5$ ,  $\eta^2 = 0.04$ ]. However, adding phase noise changed the category selectivity of the N170 as suggested by the significant category  $\times$  coherence interaction [ $F(1.57, 18.88) = 3.94$ ,  $p = 0.03$ ,  $\eta^2 = 0.25$ ].

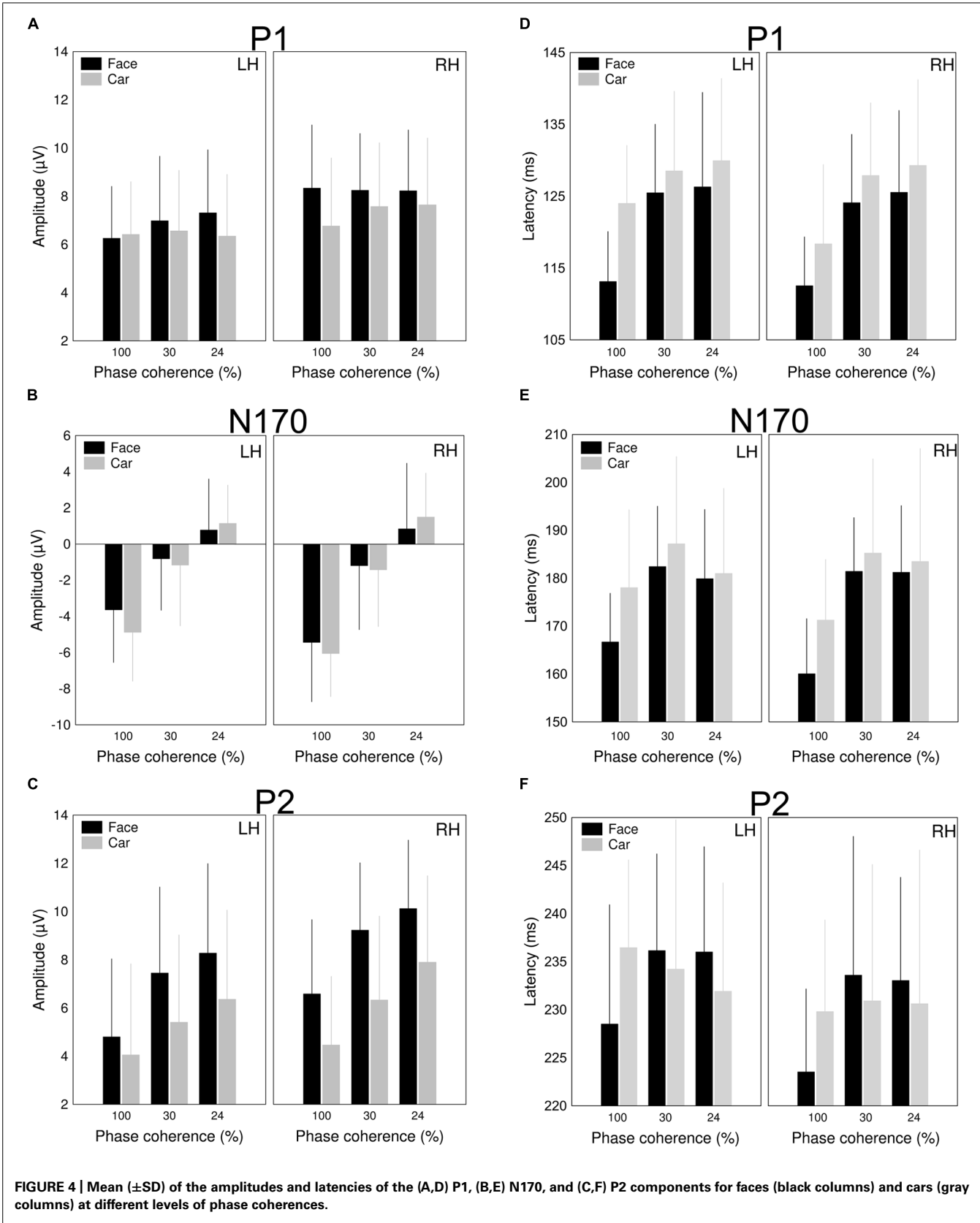
As for the N170 latency, a strong tendency of category dependence was found, suggesting that face stimuli evoked an N170 component earlier than cars [main effect of category:  $F(1, 12) = 4.32$ ,  $p = 0.06$ ,  $\eta^2 = 0.26$ ]. The N170 was delayed by adding noise to the stimulus [main effect of coherence:  $F(1.06, 12.77) = 7.82$ ,  $p = 0.0024$ ,  $\eta^2 = 0.39$ , *post hoc* LSD: 100 vs. 30 and 24%:  $p = 0.005$  for both comparisons]. In addition, a hemispheric asymmetry was also found in the noise-induced modulation of the N170 latencies [interaction between hemisphere and coherence:  $F(1.38, 16.59) = 4.8$ ,  $p = 0.0018$ ,  $\eta^2 = 0.29$ ], which was due to shorter latencies for noise absent stimuli over the RH (LSD:  $p < 0.01$ ), but similar latencies of the RH and LH for the other two noise conditions (LSD:  $ps > 0.34$ ; **Figure 4E**).

### P2

Supporting prior results (Philiastides et al., 2006; Nagy et al., 2009; Bankó et al., 2011), phase noise enhanced the amplitude of the P2 gradually [main effect of coherence:  $F(1.05, 12.6) = 25.06$ ,  $p < 0.0001$ ,  $\eta^2 = 0.68$ ]. Moreover, significantly larger P2 amplitudes were observed for face stimuli when compared to cars [main effect of category:  $F(1, 12) = 40.17$ ,  $p < 0.0001$ ,  $\eta^2 = 0.77$ , **Figure 4C**]. This effect was more pronounced in the right hemisphere, as suggested by the significant interaction between hemisphere and category [ $F(1, 12) = 6.17$ ,  $p = 0.03$ ,  $\eta^2 = 0.34$ ]. It is worth noting, however, that the category selectivity of the component was not altered by the amount of altered phase coherency [interaction between category and coherence:  $F(1.25, 15.02) = 1.6$ ,  $p = 0.22$ ,  $\eta^2 = 0.12$ ]. Finally, the P2 component also showed a strong tendency toward a RH dominance [main effect of hemisphere:  $F(1, 12) = 4.36$ ,  $p = 0.059$ ,  $\eta^2 = 0.27$ ]. No significant effects and interactions were observed on the P2 latency values (**Figure 4F**).

### The effect of stimulus ambiguity

Recent results suggest that stimulus ambiguity plays a role in determining the susceptibility of the N170 to stimulus adaptation (Walther et al., 2013). In order to test the effect of stimulus ambiguity and its noise dependence, we compared the early ERP components for the endpoints of morph continua (oldest and youngest stimuli) and for the most ambiguous (i.e., middle-aged) stimulus groups (see Materials and Methods). The first ERP component reflecting stimulus ambiguity was the N170: its amplitude was larger for middle-aged stimuli, as suggested by the main effect of age [ $F(1.9, 12.75) = 10.13$ ,  $p = 0.0006$ ,  $\eta^2 = 0.46$ ; *post hoc* LSD tests: old vs. middle-aged:  $p = 0.0004$ , young vs. middle-aged:  $p = 0.001$  but young vs. old  $p = 0.66$ ,



respectively]. N170 also had a RH dominance for middle-aged and young stimuli but not for old ones (significant hemisphere  $\times$  age interaction:  $F(1.5,18.1) = 12.53$ ,  $p = 0.0002$ ,  $\eta^2 = 0.51$ , Fisher's LSD tests:  $p < 0.0001$  for both young and middle-aged stimuli and  $p = 0.31$  for old stimuli, respectively). Interestingly, larger N170 amplitudes were measured for old cars when compared to faces, as suggested by the significant interaction between category and age [ $F(1.84,22.14) = 16.98$ ,  $p < 0.0001$ ,  $\eta^2 = 0.59$ ; Fisher's LSD:  $p < 0.0001$  for old stimuli but not for middle-aged or young stimuli:  $ps > 0.13$ ]. This effect was more pronounced in the RH [three-way interaction among category, hemisphere and age:  $F(1.65,19.78) = 8.13$ ,  $p = 0.002$ ,  $\eta^2 = 0.4$ ; **Figure 5**]. Another interesting result is that for noisy stimuli, the younger the faces were, the more pronounced the category effect was, as suggested by the significant category  $\times$  coherence  $\times$  age interaction [ $F(2.35,28.23) = 10.12$ ,  $p < 0.0001$ ,  $\eta^2 = 0.46$ ].

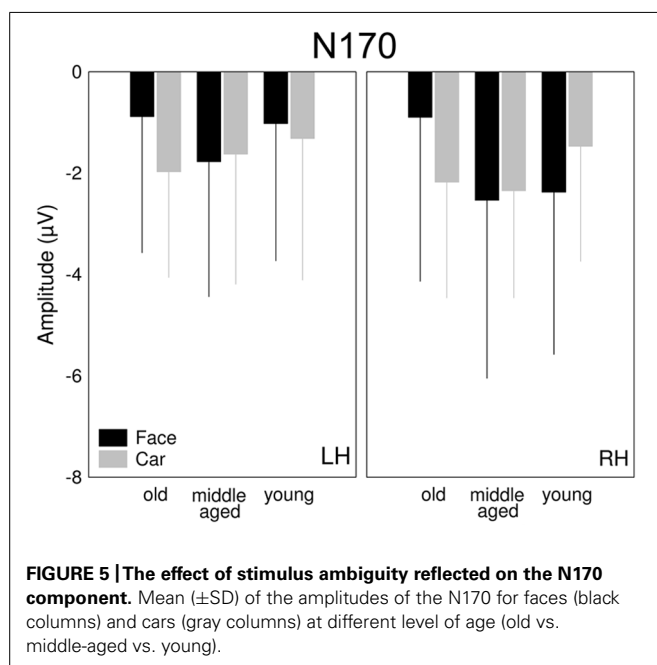
We observed larger P2 components for face than for car stimuli at every level of stimulus ambiguity, but this effect was the most pronounced for the old stimuli and weaker for the young ones (significant category  $\times$  age interaction:  $F(1.52,18.2) = 8.01$ ,  $p = 0.002$ ,  $\eta^2 = 0.4$ ). No other effect of stimulus ambiguity was found.

## DISCUSSION

The goal of the present study was to test whether adding phase noise to a stimulus affects the neural processing of complex object stimuli in a category-specific manner by recording ERPs for faces and cars at different levels of phase coherences. Several previous ERP studies applied different types of noise to manipulate the difficulty of decisions about faces (Bentin et al., 1996; McKone et al., 2001; Jemel et al., 2003; Wild and Busey, 2004; Rousselet et al., 2008b; Bankó et al., 2011). They found that adding noise to faces (or reducing their phase coherence) affects the P1 – N170 – P2

ERP complex. In the case of the face-specific N170, it was found that phase noise reduces its amplitude dramatically (Jemel et al., 2003; Nagy et al., 2009; Bankó et al., 2011) and also prolongs its latency. In addition to the changes observed in the N170 amplitude, different types of noise manipulations made the behavioral task more difficult *per se* and this difficulty was linked to the P2 ERP component (Philiastides et al., 2006; Heekeren et al., 2008): the amplitude of this component was enhanced parallel to the difficulty of the task. Later, however, it was shown that the noise-induced modulation of the P2 reflects increased visual cortical processing demands instead of task difficulty *per se* (Bankó et al., 2011). Although the effect of phase noise on the electrophysiological correlates of face perception has been investigated extensively, the question whether noise-induced modulation of these components is specific to the category of faces has so far remained unanswered. The present results suggest that the early P1 component shows a category-dependent modulation of phase coherence.

The results of the electrophysiological recordings suggest that the first stage where category-dependent phase noise-induced modulation can be observed is the level of the early P1 component. In the noise-absent conditions, faces elicited larger P1 amplitudes when compared with cars in the RH, while no such category-specific effects were found in the LH (for similar results see Itier and Taylor, 2004). P1 is usually referred to as an early indicator of the endogenous processing of visual stimuli, and it is especially linked to spatial processing (Mangun, 1995). Recently, however, it has been shown that P1 reflects more than simply the low-level features such as contrast or luminance of the stimuli, it also indexes an early stage of visual processing, being sensitive to stimulus category such as faces (Taylor, 2002). As noted by Itier and Taylor (2002), P1 could reflect the holistic processing of a face as a face, whereas the later N170 component would reflect facial configurations. Adding noise to a face causes enhanced P1 in some studies (Schneider et al., 2007; Rousselet et al., 2008b; Nagy et al., 2009; Bankó et al., 2011, 2013), while others suggest that P1 is unaffected by such changes (Jemel et al., 2003; Wild and Busey, 2004; Horovitz et al., 2004). Interestingly, in both cases, it has been suggested that P1 is not involved in any aspect of face-processing, but it is rather involved in the sensory analyses of the images, irrespective of their content (Jemel et al., 2003). In the present study, the noise-induced modulation of the P1 showed category-sensitivity in a hemisphere-specific manner. Adding noise enhanced P1 amplitudes for cars over the RH but it had no effect over the LH, and vice versa; enhanced P1 values were observed for faces in the LH but not in the RH. These results suggest that the category-specificity of the noise-induced modulation of the ERP appear very early, that is, already at 100 ms after stimulus onset. Although we have equated all stimuli in luminance and matched their histograms, we did not equate the spectral content of faces and cars since larger amount of higher spatial frequency information is caused by wrinkling and reduced skin elasticity in the case of face stimuli. Since facial age decision is mainly based on this information (e.g., George and Hole, 2000) we did not equate the spectral content of images. This, however, raises the possibility that the category-specificity of the noise-induced modulation of the early P1 component is merely



the effect of the different spectral content of the original stimuli. Indeed, several studies indicate that there are differences in sensitivity to the specific spatial frequencies both between different visual areas and between the two hemispheres (Ivry and Robertson, 1998). Our results, however, show that the category-specificity of the noise-induced modulation of the P1 is unaffected by the perceived age of the stimuli. Therefore, the amount of wrinkling that leads higher spatial frequency content in case of older faces does not modulate the results. In terms of hemispheric differences in sensitivity to specific spectral content, Sergent (1982) argued that the left hemisphere is more adept in processing high-frequency information, whereas the right hemisphere is more efficient in processing low-frequency information. This differential frequency processing account was supported by studies using tasks such as spatial frequency discrimination (Proverbio et al., 2002) and identification (Kitterle et al., 1990) or face recognition (Keenan et al., 1989). These results, however, would predict that 100% phase coherent faces with a larger amount of higher spatial frequency content would enhance the amplitude of the P1 component over the left hemisphere and adding phase noise would not affect P1 over the right hemisphere. Vice versa, 100% phase coherent cars with relatively lower spatial frequency content would enhance the P1 amplitude on the right hemisphere and adding phase noise would not affect this value when compared with the left hemisphere. However, our results show the complete opposite effect, suggesting that low-level features are not able to explain the described category dependence of P1. It is worth noting, however, that in a recent study, Motoyoshi et al. (2007) drew attention to other image-statistics that are sensitive to asymmetries in dark and light and can also affect the low-level properties of an image. Although we cannot exclude the possibility that these properties affect our results, it is unlikely that low-level differences between cars and faces are responsible for the results regarding the hemispheric asymmetries of the category-specific phase-coherence dependence of the P1 in the current study. It is also well known that there are hemispheric asymmetries in the processing of local versus global information processing. A left hemisphere advantage for responses to local features and a right hemisphere dominance for responses to global features was found in most studies (Weismann and Woldorff, 2005; Flevakis et al., 2010; Hsiao et al., 2013). Several lines of evidence suggest (e.g., the face inversion effect, the Thatcher illusion, or the composite face effect) that faces are not perceived as collections of isolated parts, but rather as holistic configurations (Yin, 1969; Thompson, 1980; Young et al., 1987). Most of the electrophysiological research studying the N170 emphasizes the specificity of the component to the structural encoding step of face processing (e.g., Bentin and Deouell, 2000; Eimer, 2000a,b). Other studies highlight the right hemisphere advantage of the component for manipulations of configural facial information, whereas the N170 in the left hemisphere is sensitive to the manipulations of featural facial information (Rossion et al., 1999; Scott and Nelson, 2006; Jacques and Rossion, 2007). This finding of different hemispheric specializations is consistent with evidence from neuroimaging studies. For example, in a PET study, Rossion et al. (2000b) have found hemispheric asymmetries for whole-based and part-based processing of faces in the fusiform gyrus in the sense that more

pronounced right fusiform activation was observed for whole faces than face parts whereas this effect was reversed in the homologous left hemisphere brain region. fMRI studies have identified a number of areas – such as the fusiform face area (FFA; Kanwisher et al., 1997) and the occipital face area (OFA; Gauthier et al., 2000) in the extrastriate visual cortex – that respond more to pictures of faces than other objects, with a strong right hemisphere dominance (McCarthy et al., 1997; Haxby et al., 1999; Rossion et al., 2003). Presumably this right hemisphere dominance is reflected in the early P1 ERP component as well. Taken together with our findings on the P1 component, we can hypothesize that the activation of the right FFA is more robust to the amount of phase noise in the case of face stimuli. In other words, it suggests that while adding phase noise to faces alters rather featural but not configural information, the right hemisphere will be unaffected by this image manipulation. Although in a source localization study investigating the early stages of face processing, Herrmann et al. (2005b) have shown that the first step of cortical face processing (~100 ms after stimulus presentation) is localized in the fusiform gyrus, further studies are needed to clarify the sensitivity of the FFA to image manipulations such as the effect of phase noise.

The electrophysiological results of the current study confirmed the classical noise-induced effects reflected in the N170 and P2 components (Nagy et al., 2009; Bankó et al., 2011): the N170 amplitude decreased for higher levels of phase noise in a stepwise manner (Jemel et al., 2003). The gradual decrease of the N170 as the faces and cars became more and more noisy can be accounted for by the sensitivity of the component to the visibility of the stimuli embedded in different amounts of noise. It can also be due to increased attentional resources as the amount of added phase noise reduced the coherence of the stimuli. The fact that the observed significant three-way category  $\times$  coherence  $\times$  hemisphere interaction measured on the P1 lost its hemispheric asymmetry in the N170 time window is suggestive of the involvement of additional neural mechanisms. Schneider et al. (2007) have shown that noise affects the neural correlates of upright and inverted faces differently. Many studies suggest that inversion results in faces being processed by a piecemeal, feature-by-feature strategy (Rossion et al., 2000a; Barton et al., 2001), more similar to non-face objects (Haxby et al., 1999; Rossion et al., 2000a; Rosburg et al., 2010; Kloth et al., 2013). As complex, non-face object stimuli such as cars are also processed in a feature-based manner, the category  $\times$  coherence interaction observed in the N170 component is rather due to the effect of stimulus configuration on processing levels. The fact that N170 was similar in amplitude for 100% phase coherent car and face images suggests that individual exemplars of objects that are visually similar to faces and have homogeneous feature configurations can elicit comparable N170 responses (for similar stimulus comparisons and results see Kloth et al., 2013). It is worth noting, however, that these results do not suggest that similar encoding takes place for cars and faces, even when they are characterized by a similar, face-like configuration (Kloth et al., 2013). On the other hand, our results also confirm the classical noise-induced effects on the later P2 component as well (Nagy et al., 2009; Bankó et al., 2011). More positive peaks were observed



for faces when compared to cars, especially in the RH, and gradually increased P2 components were measured parallel to the amount of added noise. In previous studies, the noise induced effect reflected in the later P2 component could be explained by two factors – adding noise to the stimulus increases the visual cortical processing demands (Bankó et al., 2011, 2013), or it results in enhanced responses of the neural populations representing stimulus uncertainty (Bach and Dolan, 2012). Since no significant category  $\times$  coherence interaction was observed on the P2 component the results of the current study could not exclude either explanation.

In summary, in this electrophysiological study we explicitly compared the noise-dependence of face and non-face stimuli and we have found that the neural processing of different high-level categories diverge at a very early stage of stimulus processing, starting in the P1 time window.

## AUTHOR CONTRIBUTIONS

Designed the experiment: Kornél Németh, Gyula Kovács, Márta Zimmer; data acquisition: Kornél Németh, Petra Kovács, Pál Vakli; data analyses: Kornél Németh, Petra Kovács, Pál Vakli, Gyula Kovács, Márta Zimmer; interpretation of the data: Gyula Kovács, Márta Zimmer; provided materials: Kornél Németh, Petra Kovács, Pál Vakli, Márta Zimmer; wrote the article: Gyula Kovács, Márta Zimmer; proofed/revised the article: Kornél Németh, Petra Kovács, Pál Vakli, Gyula Kovács, Márta Zimmer.

## ACKNOWLEDGMENTS

This work was supported by the Hungarian Scientific Research Fund (OTKA) PD 101499 (Márta Zimmer), by the Deutsche Forschungsgemeinschaft (KO 3918/1-1; Gyula Kovács) and by the National Development Agency (TÁMOP; TÁMOP-4.2.2/B-10/1-2010-0009; Petra Kovács).

## REFERENCES

- Bach, D. R., and Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* 13, 572–586. doi: 10.1038/nrn3289
- Bankó, E. M., Gál, V., Körtvélyes, J., Kovács, G., and Vidnyánszky, Z. (2011). Dissociating the effect of noise on sensory processing and overall decision difficulty. *J. Neurosci.* 31, 2663–2674. doi: 10.1523/JNEUROSCI.2725-10.2011
- Bankó, E. M., Gál, V., and Vidnyánszky, Z. (2009). Flawless visual short-term memory for facial emotional expressions. *J. Vis.* 9, 12.1–12.13. doi: 10.1167/9.1.12
- Bankó, E. M., Körtvélyes, J., Weiss, B., and Vidnyánszky, Z. (2013). How the visual cortex handles stimulus noise: insights from amblyopia. *PLoS ONE* 8:e66583. doi: 10.1371/journal.pone.0066583
- Barton, J., Keenan, J., and Bass, T. (2001). Discrimination of spatial relations and features in faces: effects of inversion and viewing duration. *Br. J. Psychol.* 92, 527–549. doi: 10.1348/000712601162329
- Bentin, S., Allison, T., Puce, A., Perez, E., and McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *J. Cogn. Neurosci.* 8, 551–565. doi: 10.1162/jocn.1996.8.6.551
- Bentin, S., and Deouell, L. Y. (2000). Structural encoding and identification in face processing: erp evidence for separate mechanisms. *Cogn. Neuropsychol.* 17, 35–55. doi: 10.1080/026432900380472
- Boutsen, L., Humphreys, G. W., Praamstra, P., and Warbrick, T. (2006). Comparing neural correlates of configural processing in faces and objects: an ERP study of the Thatcher illusion. *Neuroimage* 32, 352–367. doi: 10.1016/j.neuroimage.2006.03.023
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Chatrian, G. E., Wirth, A. L., Edwards, K. H., Turella, G. S., Kaufman, M. A., and Snyder, J. M. (1985). Cochlear summing potential to broadband clicks detected from the human external auditory meatus. A study of subjects with normal hearing for age. *Ear Hear.* 6, 130–138. doi: 10.1097/00003446-198505000-00002
- Coleman, S. R., and Grover, R. (2006). The anatomy of aging face: volume loss and change in 3-dimensional topography. *Aesthet. Surg. J.* 26, S4–S9. doi: 10.1016/j.asj.2005.09.012
- Costen, N. P., Parker, D. M., and Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception* 23, 129–146. doi: 10.1068/p230129
- Curran, T., Tucker, D. M., Kutas, M., and Posner, M. I. (1993). Topography of the N400: brain electrical activity reflecting semantic expectancy. *Electroencephalogr. Clin. Neurophysiol.* 88, 188–209. doi: 10.1016/0168-5597(93)90004-9
- Dakin, S. C., Hess, R. F., Ledgeway, T., and Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Curr. Biol.* 12, R476–R477; author reply R478.
- Dering, B., Martin, C. D., Moro, S., Pegna, A. J., and Thierry, G. (2011). Face-sensitive processes one hundred milliseconds after picture onset. *Front. Hum. Neurosci.* 5:93. doi: 10.3389/fnhum.2011.00093
- Eimer, M. (2000a). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport* 11, 2319–2324. doi: 10.1097/00001756-200007140-00050
- Eimer, M. (2000b). Effects of face inversion on the structural encoding and recognition of faces: evidence from event-related brain potentials. *Cogn. Brain Res.* 10, 145–158. doi: 10.1016/S0926-6410(00)00038-0
- Eimer, M. (2011). The face-sensitivity of the n170 component. *Front. Hum. Neurosci.* 5:119. doi: 10.3389/fnhum.2011.00119
- Flevaris, A. V., Bentin, S., and Robertson, L. C. (2010). Local or global? Attentional selection of spatial frequencies binds shapes to hierarchical levels. *Psychol. Sci.* 21, 424–431. doi: 10.1177/0956797609359909
- Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). The fusiform “face area” is part of a network that processes faces at the individual level. *J. Cogn. Neurosci.* 12, 495–504. doi: 10.1162/089892900562165
- George, N., Jemel, B., Fiori, N., and Renault, B. (1997). Face and shape repetition effects in humans: a spatio-temporal ERP study. *Neuroreport* 8, 1417–1423. doi: 10.1097/00001756-199704140-00019
- George, P. A., and Hole, G. J. (2000). The role of spatial and surface cues in the age-processing of unfamiliar faces. *Vis. Cogn.* 7, 485–509. doi: 10.1080/135062800394621
- Gold, J., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature* 402, 176–178. doi: 10.1038/46027
- Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr. Clin. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9
- Halit, H., De Haan, M., and Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport* 11, 1871–1875. doi: 10.1097/00001756-200006260-00014
- Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., and Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* 22, 189–199. doi: 10.1016/S0896-6273(00)80690-X
- Heekeren, H. R., Marrett, S., and Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479. doi: 10.1038/nrn2374
- Herrmann, M. J., Ehlis, A. C., Ellgring, H., and Fallgatter, A. J. (2005a). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *J. Neural. Trans.* 112, 1073–1081. doi: 10.1007/s00702-004-0250-8
- Herrmann, M. J., Ehlis, A. C., Muehlberger, A., and Fallgatter, A. J. (2005b). Source localization of early stages of face processing. *Brain Topogr.* 18, 77–85. doi: 10.1007/s10548-005-0277-7
- Horowitz, S. G., Rossion, B., Skudlarski, P., and Gore, J. C. (2004). Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *Neuroimage* 22, 1587–1595. doi: 10.1016/j.neuroimage.2004.04.018

- Hsiao, J. H., Cipollini, B., and Cottrell, G. W. (2013). Hemispheric asymmetry in perception: a differential encoding account. *J. Cogn. Neurosci.* 25, 998–1007. doi: 10.1162/jocn\_a\_00377
- Itier, R. J., and Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuroimage* 15, 353–372. doi: 10.1006/nimg.2001.0982
- Itier, R. J., and Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cereb. Cortex* 14, 132–142. doi: 10.1093/cercor/bhg111
- Ivry, R., and Robertson, L. C. (1998). *The Two Sides of Perception*. Cambridge, MA: MIT Press.
- Jacques, C., and Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *Neuroimage* 36, 863–876. doi: 10.1016/j.neuroimage.2007.04.016
- Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., and Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuroreport* 14, 2035–2039. doi: 10.1097/01.wnr.0000092465.31470.2f
- Johannes, S., Munte, T. F., Heinze, H. J., and Mangun, G. R. (1995). Luminance and spatial attention effects on early visual processing. *Cogn. Brain Res.* 2, 189–205. doi: 10.1016/0926-6410(95)90008-X
- Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- Keenan, P. A., Whitman, P. D., and Pepe, J. (1989). Hemispheric asymmetry in the processing of high and low spatial frequencies: a facial recognition task. *Brain Cogn.* 11, 229–237. doi: 10.1016/0278-2626(89)
- Kitterle, F. L., Christman, S., and Hellige, J. B. (1990). Hemispheric differences are found in the identification, but not the detection, of low versus high spatial frequencies. *Percept. Psychophysiol.* 48, 297–306. doi: 10.3758/BF03211607
- Kloth, N., Itier, R. J., and Schweinberger, S. R. (2013). Combined effects of inversion and feature removal on N170 responses elicited by faces and car fronts. *Brain Cogn.* 81, 321–328. doi: 10.1016/j.bandc.2013.01.002
- Kovács, G., Zimmer, M., Bankó, E., Harza, I., Antal, A., and Vidnyánszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cereb. Cortex* 16, 742–753. doi: 10.1093/cercor/bhj020
- Kovács, G., Zimmer, M., Harza, I., Antal, A., and Vidnyánszky, Z. (2005). Position-specificity of facial adaptation. *Neuroreport* 16, 1945–1949. doi: 10.1097/01.wnr.0000187635.76127.bc
- Kovács, G., Zimmer, M., Harza, I., and Vidnyánszky, Z. (2007). Adaptation duration affects the spatial selectivity of facial aftereffects. *Vision Res.* 47, 3141–3149. doi: 10.1016/j.visres.2007.08.019
- Latinus, M., and Taylor, M. J. (2006). Face processing stages: impact of difficulty and the separation of effects. *Brain Res.* 1123, 179–187. doi: 10.1016/j.brainres.2006.09.031
- Lee, B., and Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Res.* 36, 2159–2166. doi: 10.1016/0042-6989(95)00271-5
- Liu, J., Harris, A., and Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nat. Neurosci.* 5, 910–916. doi: 10.1038/nn909
- Mangun, G. R. (1995). Neural mechanisms of visual selective attention. *Psychophysiology* 32, 4–18. doi: 10.1111/j.1469-8986.1995.tb03400.x
- McCarthy, G., Puce, A., Gore, J. C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9, 605–610. doi: 10.1162/jocn.1997.9.5.605
- McKone, E., Martini, P., and Nakayama, K. (2001). Categorical perception of face identity in noise isolates configural processing. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 573–599. doi: 10.1037/0096-1523.27.3.573
- Mercure, E., Dick, F., and Johnson, M. H. (2008). Featural and configural face processing differentially modulate ERP components. *Brain Res.* 1239, 162–170. doi: 10.1016/j.brainres.2008.07.098
- Milivojevic, B., Clapp, W. C., Johnson, B. W., and Corballis, M. C. (2003). Turn that frown upside down: ERP effects of thatcherization of misorientated faces. *Psychophysiology* 40, 967–978. doi: 10.1111/1469-8986.00115
- Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Instrum. Comput.* 36, 630–633. doi: 10.3758/BF03206543
- Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature* 447, 206–209. doi: 10.1038/nature05724
- Nagy, K., Zimmer, M., Liu, W., and Kovács, G. (2009). The sensitivity of face specific ERP components to the nature of stimulus noise. *Learn. Percept.* 1, 183–197. doi: 10.1556/LP.1.2009.2.3
- Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Res.* 39, 3824–3833. doi: 10.1016/S0042-6989(99)00096-6
- Olman, C., Boyaci, H., Fang, F., and Doerschner, K. (2008). V1 responses to different types of luminance histogram contrast [Abstract]. *J. Vis.* 8, 345a. doi: 10.1167/8.6.345
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Philastides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* 26, 8965–8975. doi: 10.1523/JNEUROSCI.1655-06.2006
- Proverbio, A. M., Zani, A., and Avella, C. (2002). Hemispheric asymmetries for spatial frequency discrimination in a selective attention task. *Brain Cogn.* 34, 311–320. doi: 10.1006/brcg.1997.0901
- Rosburg, T., Ludwig, E., Dumpelmann, M., Alba-Ferrara, L., Urbach, H., and Elger, C. E. (2010). The effect of face inversion on intracranial and scalp recordings of event-related potentials. *Psychophysiology* 47, 147–157. doi: 10.1111/j.1469-8986.2009.00881.x
- Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., and Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain* 126, 2381–2395. doi: 10.1093/brain/awg241
- Rossion, B., Delvenne, J. F., Debatiste, D., Goffaux, V., Bruyer, R., Crommelinck, M., et al. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biol. Psychol.* 50, 173–189. doi: 10.1016/S0301-0511(99)00013-7
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (2000a). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport* 11, 69–74. doi: 10.1097/00001756-200001170-00014
- Rossion, B., Dricot, L., Devolder, A., Bodart, J. M., Crommelinck, M., De Gelder, B., et al. (2000b). Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 12, 793–802. doi: 10.1162/08992900562606
- Rossion, B., and Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage* 39, 1959–1979. doi: 10.1016/j.neuroimage.2007.10.011
- Rossion, B., and Jacques, C. (2011). “The N170 : understanding the time-course of face perception in the human brain,” in *The Oxford Handbook of ERP Components*, eds S. Luck and E. Kappenman. (Oxford: Oxford University Press), 115–141.
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2008a). Time course and robustness of ERP object and face differences. *J. Vis.* 8, 1–18. doi: 10.1167/8.12.3
- Rousselet, G. A., Pernet, C. R., Bennett, P. J., and Sekuler, A. B. (2008b). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neurosci.* 9:98. doi: 10.1186/1471-2202-9-98
- Rousselet, G. A., Husk, J. S., Bennett, P. J., and Sekuler, A. B. (2007). Single-trial EEG dynamics of object and face visual processing. *Neuroimage* 36, 843–862. doi: 10.1016/j.neuroimage.2007.02.052
- Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A. N., Heller, A. C., Ross, I. B., et al. (2011). Single-unit responses selective for whole faces in the human amygdala. *Curr. Biol.* 21, 1654–1660. doi: 10.1016/j.cub.2011.08.035
- Schneider, B. L., Delong, J. E., and Busey, T. A. (2007). Added noise affects the neural correlates of upright and inverted faces differently. *J. Vis.* 7, 4. doi: 10.1167/7.4.4
- Schyns, P. G., Gosselin, F., and Smith, M. (2009). Information processing algorithms in the brain. *Trends Cogn. Sci.* 13, 20–26. doi: 10.1016/j.tics.2008.09.008



- Schyns, P. G., Jentzsch, I., Johnson, M., Schweinberger, S. R., and Gosselin, F. (2003). A principled method for determining the functionality of brain responses. *Neuroreport* 14, 1665–1669. doi: 10.1097/00001756-200309150-00002
- Schyns, P. G., Petro, L. S., and Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Curr. Biol.* 17, 181580–1585. doi: 10.1016/j.cub.2007.08.048
- Schweinberger, S. R., Huddy, V., and Burton, A. M. (2004). N250r: a face-selective brain response to stimulus repetitions. *Neuroreport* 15, 1501–1505. doi: 10.1097/01.wnr.0000131675.00319.42
- Scott, L. S., and Nelson, C. A. (2006). Featural and configural face processing in adults and infants: a behavioral and electrophysiological investigation. *Perception* 35, 1107–1128. doi: 10.1068/p5493
- Seeck, M., Michel, C. M., Mainwaring, N., Cosgrove, R., Blume, H., Ives, J., et al. (1997). Evidence for rapid face recognition from human scalp and intracranial electrodes. *Neuroreport* 8, 2749–2754. doi: 10.1097/00001756-199708180-00021
- Sergent, J. (1982). The cerebral balance of power: confrontation or cooperation? *J. Exp. Psychol. Hum. Percept. Perform.* 8, 253–272. doi: 10.1037/0096-1523.8.2.253
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychol. Sci.* 15, 753–761. doi: 10.1111/j.0956-7976.2004.00752.x
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5626–5631. doi: 10.1073/pnas.0508972103
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2007). From a face to its category via a few information processing states in the brain. *Neuroimage* 37, 974–984. doi: 10.1016/j.neuroimage.2007.05.030
- Smith, F. W., Muckli, L., Brennan, D., Pernet, C., Smith, M. L., Belin, P., et al. (2008). Classification images reveal the information sensitivity of brain voxels in fMRI. *Neuroimage* 40, 1643–1654. doi: 10.1016/j.neuroimage.2008.01.029
- Smith, M., Fries, P., Gosselin, F., Goebel, R., and Schyns, P. G. (2009). Inverse mapping the neuronal substrates of face categorizations. *Cereb. Cortex* 19, 2428–2438. doi: 10.1093/cercor/bhn257
- Smith, M. L., Gosselin, F., and Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Curr. Biol.* 22, 191–196. doi: 10.1016/j.cub.2011.11.061
- Stahl, J., Wiese, H., and Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: an event-related potential study. *Neuroreport* 19, 583–587. doi: 10.1097/WNR.0b013e3282f97b4d
- Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clin. Neurophysiol.* 113, 1903–1908. doi: 10.1016/S1388-2457(02)00309-7
- Thierry, G., Martin, C. D., Downing, P., and Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nat. Neurosci.* 10, 505–511. doi: 10.1038/nn1864
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception* 9, 483–484. doi: 10.1068/p090483
- Tjan, B. S., Lestou, V., and Kourtzi, Z. (2006). Uncertainty and invariance in the human visual cortex. *J. Neurophysiol.* 96, 1556–1568. doi: 10.1152/jn.01367.2005
- Tucker, D. M., Liotti, M., Potts, G. F., Russell, G. S., and Posner, M. I. (1994). Spatiotemporal analysis of brain electrical fields. *Hum. Brain Mapp.* 1, 134–152. doi: 10.1002/hbm.460010206
- Walther, C., Schweinberger, S. R., and Kovács, G. (2013). Adaptor identity modulates adaptation effects in familiar face identification and their neural correlates. *PLoS ONE* 8:e70525. doi: 10.1371/journal.pone.0070525
- Weismann, D. H., and Woldorff, M. G. (2005). Hemispheric asymmetries for different components of global/local attention occur in distinct temporo-parietal loci. *Cereb. Cortex* 15, 870–876. doi: 10.1093/cercor/bhh187
- Wichmann, F. A., and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313. doi: 10.3758/BF03194544
- Wiese, H., Stahl, J., and Schweinberger, S. R. (2009). Configural processing of other-race faces is delayed but not decreased. *Biol. Psychol.* 81, 103–109. doi: 10.1016/j.biopsycho.2009.03.002
- Wild, H. A., and Busey, T. A. (2004). Seeing faces in the noise: stochastic activity in perceptual regions of the brain may influence the perception of ambiguous stimuli. *Psychon. Bull. Rev.* 11, 475–481. doi: 10.3758/BF03196598
- Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behav. Res. Methods* 42, 671–684. doi: 10.3758/BRM.42.3.671
- Yin, R. K. (1969). Looking at upside-down faces. *J. Exp. Psychol.* 81, 141–145. doi: 10.1037/h0027474
- Young, A., Hellawell, D., and Hay, D. C. (1987). Configurational information in face perception. *Perception* 16, 737–759. doi: 10.1068/p160747

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 November 2013; accepted: 07 April 2014; published online: 24 April 2014.

Citation: Németh K, Kovács P, Vakli P, Kovács G and Zimmer M (2014) Phase noise reveals early category-specific modulation of the event-related potentials. *Front. Psychol.* 5:367. doi: 10.3389/fpsyg.2014.00367

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Németh, Kovács, Vakli, Kovács and Zimmer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Observer efficiency in free-localization tasks with correlated noise

Craig K. Abbey\* and Miguel P. Eckstein

Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA

**Edited by:**

Rémy Allard, Université Pierre et Marie Curie, France

**Reviewed by:**

Markus Lappe, Universität Münster, Germany

Peter Neri, University of Aberdeen, UK

**\*Correspondence:**

Craig K. Abbey, Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA 93106, USA  
e-mail: [abbey@psych.ucsb.edu](mailto:abbey@psych.ucsb.edu)

The efficiency of visual tasks involving localization has traditionally been evaluated using forced choice experiments that capitalize on independence across locations to simplify the performance of the ideal observer. However, developments in ideal observer analysis have shown how an ideal observer can be defined for free-localization tasks, where a target can appear anywhere in a defined search region and subjects respond by localizing the target. Since these tasks are representative of many real-world search tasks, it is of interest to evaluate the efficiency of observer performance in them. The central question of this work is whether humans are able to effectively use the information in a free-localization task relative to a similar task where target location is fixed. We use a yes-no detection task at a cued location as the reference for this comparison. Each of the tasks is evaluated using a Gaussian target profile embedded in four different Gaussian noise backgrounds having power-law noise power spectra with exponents ranging from 0 to 3. The free localization task had a square  $6.7^\circ$  search region. We report on two follow-up studies investigating efficiency in a detect-and-localize task, and the effect of processing the white-noise backgrounds. In the fixed-location detection task, we find average observer efficiency ranges from 35 to 59% for the different noise backgrounds. Observer efficiency improves dramatically in the tasks involving localization, ranging from 63 to 82% in the forced localization tasks and from 78 to 92% in the detect-and-localize tasks. Performance in white noise, the lowest efficiency condition, was improved by filtering to give them a power-law exponent of 2. Classification images, used to examine spatial frequency weights for the tasks, show better tuning to ideal weights in the free-localization tasks. The high absolute levels of efficiency suggest that observers are well-adapted to free-localization tasks.

**Keywords:** free-localization tasks, ideal observer theory, power-law noise, observer efficiency, image statistics

## INTRODUCTION

The concept of calculation efficiency, which we refer to simply as efficiency, in the presence of image noise has been used extensively as a method for understanding visual processing since its seminal introduction by Barlow (Barlow, 1977, 1978; Barlow and Reeves, 1979). At the core of this measure is comparison with an optimal decision maker, the ideal observer, for a given task. The use of the ideal observer as yardstick for human performance implicitly controls for the relevant information present in stimuli used to perform a task. This topic has a long history in vision science, as well as areas of applied vision such as medical imaging. In the realm of vision science, there are many examples where efficiency is used to reveal the presence (or absence) of limitations and constraints in visual processing (Barlow, 1978; Barlow and Reeves, 1979; Burgess et al., 1981; Pelli, 1985; Legge et al., 1987; Geisler, 1989; Tjan et al., 1995). In imaging applications, efficiency is used to identify opportunities for image processing or other methodological changes that lead to improved performance in visual tasks (Myers et al., 1985; Wagner and Brown, 1985; Insana and Hall, 1994; Siewerdsen and Jaffray, 2000; Abbey et al., 2006).

Studies evaluating efficiency have often relied on experimental paradigms where the location of a target, if it is present, is explicitly defined through the use of location cues. Forced-choice paradigms, with two or more specified locations that serve as possible target locations, are a common choice (Burgess and Ghandeharian, 1984). These studies do involve spatial (or temporal) search, but it is a limited search that is confined to choosing between distinct, cued locations. The use of independent noise masking the target at each location makes the computation of the ideal observer considerably easier. Studies that have analyzed the ideal observer in tasks with location uncertainty on a quasi-continuous scale (i.e., limited to the pixelation of the stimulus) have generally utilized a detection or discrimination response that did not involve localizing the target (Park et al., 2005; Tjan and Nandy, 2006; Neri, 2010).

However, recent analysis by Khurd and Gindi (2005) have demonstrated how an ideal observer may be evaluated when targets can be located anywhere within a search region, and the task requires localizing targets to within a fixed distance, or more general acceptance region. This general paradigm has been used previously in medical imaging studies (Burgess et al., 2001;

Bochud et al., 2004) due to the similarity with many clinical tasks that require identifying a location in the body for further assessment. However, these studies did not have the benefit of an ideal observer. The Khurd and Gindi approach leads to the definition of an optimal decision function, from which ideal observer performance can be extracted via simulation studies, as we do below. Extensions to the theory (Khurd et al., 2010) include methods for evaluating the presence of multiple targets that are beyond the scope of this work. There has been some use of this analysis to evaluate the role of regularization in emission computed tomography (Liu et al., 2009). However, we are not aware of any use of the ideal observer for examining efficiency in more general free-localization tasks.

The main focus of this work is a comparison of fixed-location detection tasks—where a single target location is well cued—to free-localization tasks, where the subject must indicate the location of a target that can be anywhere in a defined search region. **Figure 1** gives examples of the stimulus displays for each task. For the detection tasks, subjects render a decision on whether a Gaussian “bump” target profile is present or not at the cued location. In the localization tasks, the subject is required to indicate the location of the target profile, which will always be present somewhere in the search region. We are interested in comparing the efficiency of human observers in these two tasks, and understanding the mechanisms that can explain differences.

We are also interested in the role of the background image statistics on this process. For this reason, we evaluate four different Gaussian image textures. These are defined by their power spectra, which are constrained to be a power-law parameterized by the power-law exponent  $\beta$ . We evaluated four different background textures, based on  $\beta$ -values ranging from 0 (white noise) to 3 (See **Figure 2** below). Natural scenes are often modeled as power-law processes with exponents that vary around  $\beta = 2$  (Burton and Moorhead, 1987; Field, 1987). Various forms of x-ray images, for breast imaging in particular, have also been

modeled as power-law processes with exponents ranging from less than 2 for computed tomography reconstructions (Metheany et al., 2008; Chen et al., 2012, 2013) to 3 or more for tomosynthesis (Engstrom et al., 2009) or projection images (Bochud et al., 1999; Burgess et al., 2001).

In addition to the main comparison of detection and localization tasks, two smaller follow-up studies were conducted to give additional insight on issues that arose from the primary study. One issue was the different nature of the two tasks, given that a yes-no type detection task requires maintaining some sort of detection criterion from trial to trial for choosing the response. The free-localization tasks do not require this. To investigate the effect of a detection criterion on free-localization tasks, we evaluated a detect-and-localize (D&L) task, in which the target profile appeared at a random location in the search region in 50% of the trials, and was not present in the other 50% of the trials. As an alternative to indicating the target location, the subjects could also respond “not present” in these experiments. In this way, the requirement of maintaining a task criterion in the detection task was matched in a task with substantial spatial uncertainty.

The second follow-up study concerned the white-noise ( $\beta = 0$ ) background condition, where lower efficiency than other power-law backgrounds was observed in both detection and localization tasks. In this case, we were interested in whether processing the images to have more favorable background statistics could improve performance. We evaluated task performance after filtering these images to have a power-law power spectrum with  $\beta = 2$ , which also modified the profile of the target.

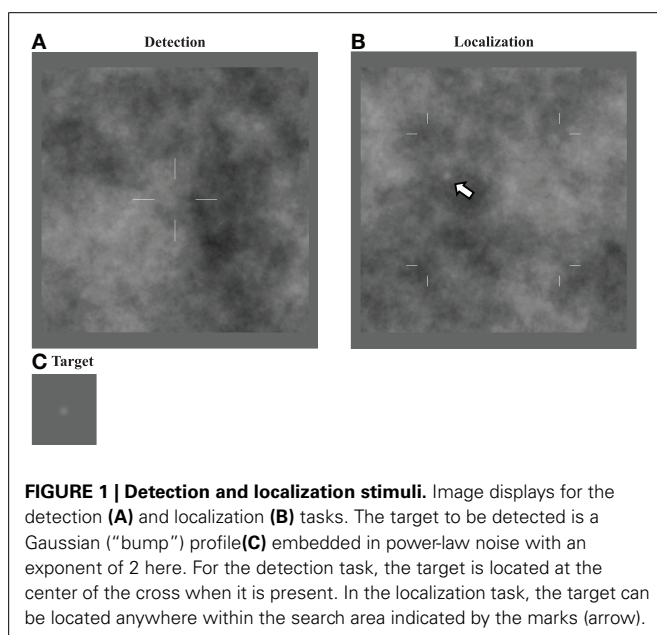
## METHODS

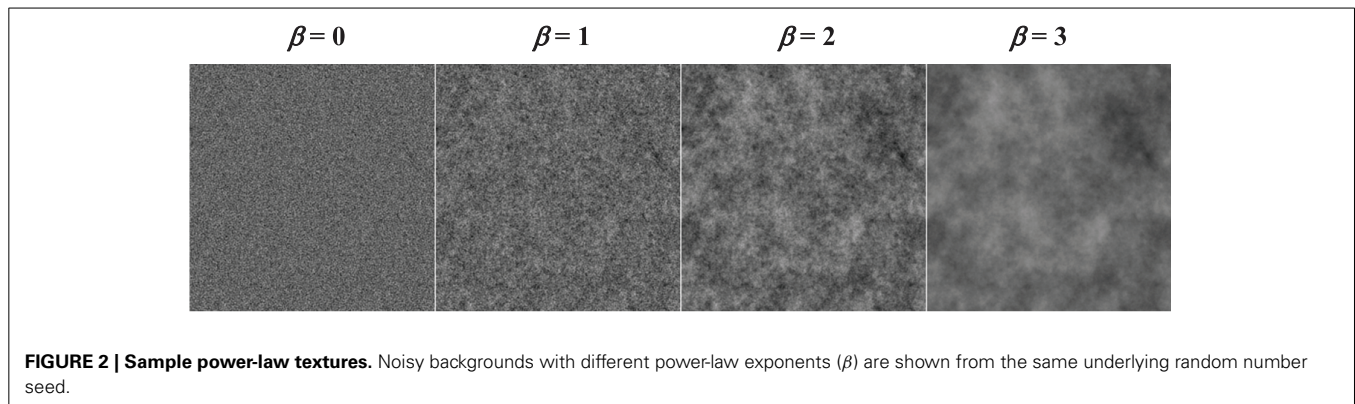
A total of 5 subjects participated in the primary comparison of efficiency between detection and forced-localization tasks. On subject (S1) was a coauthor of this work, and the other 4 were naïve to the purposes of the research and compensated for their participation. Of these, 3 subjects (S2, S4, and S5) participated in the secondary detect-and-localize experiments, and 3 subjects (S3, S4, and S5) participated in the secondary image-processing experiments.

### STIMULUS AND DISPLAY PROPERTIES

A monochrome CRT display (Imaging Systems, Minnetonka, Minnesota) with a dedicated controller (DOME, NDS Inc., San Jose, CA), was used for all experiments, which were conducted in a darkened room. The monitor was photometer-calibrated to an 8-bit linear lookup table (LUT) that ranged from 0.02–40 Cd/m<sup>2</sup>. Viewing distance of the subjects was not constrained. Subjects had normal vision or wore corrective lenses. After becoming familiar with the display procedure, and completing several sessions of experiments, measurements of each subject’s comfortable viewing distance were made. The average viewing distance used was 64 cm, with a range of 51–70 cm. This average distance was used for all subsequent calculations of visual angle. The stimuli were generated as 256 by 256 pixel images, and these were magnified by a factor of two for display, making the effective pixel size 0.052° (0.583 mm).

The experiments used a Gaussian “bump” as a target added to stationary noise with a power-law power spectrum. The spatial





standard deviation of the target was 3 pixels giving the displayed target a FWHM of  $0.37^\circ$ . A mean background level of 100 gray levels (gl) was added as well and the noise was scaled to have a pixel variance of  $400 \text{ gl}^2$ , which is equivalent to a 20% RMS contrast on the linearized display that was used for the psychophysical studies. Target contrast varied over the different experiments, as described below.

Noise backgrounds were generated by filtering white noise to achieve a power-law power spectrum in the spatial-frequency domain,  $S_\beta(f) = C_\beta/f^\beta$ , which we will identify by the power-law exponent,  $\beta$ . To avoid the singularity at  $f = 0$ , the DC component is set to the value of the first harmonic. The normalization constant,  $C_\beta$ , is set so that the RMS contrast of the background is fixed at 20%. The noise generation filter for each background condition was set to be the square-root of  $S_\beta(f)$ . Examples of the different noise textures for the four values of  $\beta$  that we used are seen in **Figure 2**.

For detection tasks using the different backgrounds, targets had a 50% probability of being present in any given trial. Subjects were informed that this was the target probability. When present, the target was always located in the center of the image with the location indicated by cross-hairs, as shown in **Figure 1**. The observer response was obtained by capturing a mouse click outside the image area. Feedback (correct/incorrect) was given after each trial. While separate performance measures were determined for target present and target absent images (hit rate and false alarm rate), for purposes of fitting psychometric functions the proportion of correct responses was also used.

For localization tasks, the target was randomly located in the central region of the image, with borders delineated by hash marks. Subjects were informed that this was the search region. The central region consisted of 128 by 128 pixels ( $6.7$  by  $6.7^\circ$ ), and thus constituted one quarter of the total image area. The large border region was chosen to minimize any edge effects as well as effects from “wrap-around” from the filtering operation. Observers responded by clicking with a mouse on their selected location. Mouse-clicks that were 5 pixels ( $0.26^\circ$ ) or less from the center of the target were considered “correct,” and subject performance was measured as the proportion of correct responses. Subjects received feedback (correct/incorrect and true target location) after each trial.

Slight modifications to the experimental protocol above were used in the two follow-up studies. For the D&L tasks, there was a 50% probability that the target was present somewhere in the search region. Subjects responded with a mouse-click on the target location to indicate target presence at that location, or by a mouse-click outside the image to indicate target not present. Subjects received feedback (correct/incorrect and target location if applicable) after each trial.

In the image processing study, white noise images (i.e.,  $\beta = 0$ ) were filtered after the target profile was added to have a  $\beta = 2$  power-law spectrum. The frequency profile of the filter was  $\sqrt{S_2(f)}$ . As a result of this filtering operation, the target was no longer a Gaussian profile, which reflects the practical reality that image processing alters the properties and appearance of both target and background. Examples of the an image before and after filtering, as well as a plot showing the effect on the target profile, are seen in **Figure 3**. Detection and localization tasks on the processed images were run as described above.

## THE IDEAL OBSERVER

Task efficiency with respect to the ideal observer is the fundamental calculation used in this work. In this section we describe how the ideal observer analysis is implemented, leading to an efficiency estimate.

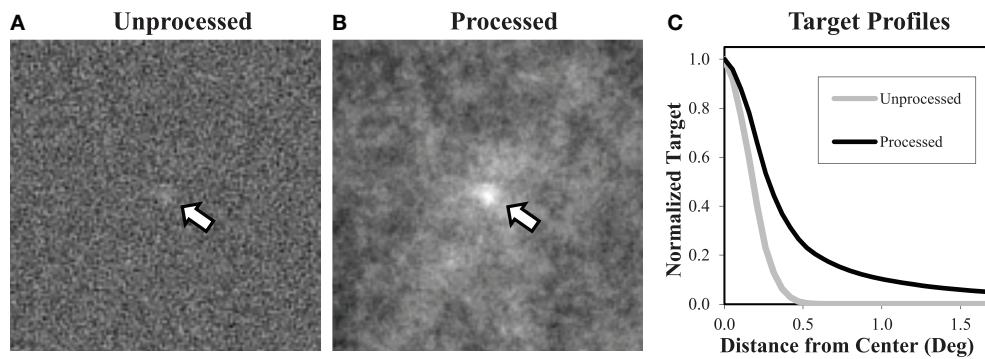
### Detection tasks

For the yes-no detection task we identify target-present images as one hypothesis (or class),  $H_1$ , and the target-absent images as the other possible hypothesis,  $H_0$ . We will refer to the images generically as  $\mathbf{g}$ , a column vector of pixel values, with the assumption that the mean background intensity of the stimuli (100 gl in our case) has been subtracted off of the pixel intensities. The Gaussian noise in the images is specified by a multivariate normal distribution (MVN) with a covariance matrix that depends on the power-law exponent of the noise texture,  $\Sigma_\beta$ . The conditional distributions of the resulting images are given by

$$\begin{aligned} p(\mathbf{g}|H_0) &= \text{MVN}(\mathbf{0}, \Sigma_\beta) \\ p(\mathbf{g}|H_1) &= \text{MVN}(\mathbf{s}, \Sigma_\beta). \end{aligned} \quad (1)$$

Under these conditions, it is well known that the ideal observer can be implemented as a weighted sum of the pixel intensities





**FIGURE 3 | Processing white noise images.** Processing white noise images (A) with the appropriate filter gives them a  $\beta = 2$  power-law background (B). This transformation also changes the target profile, giving it much longer tails (C). Target contrast (arrows) is enhanced here for display.

(Green and Swets, 1966). Let the vector  $\mathbf{w}_{\text{IO},\beta}$  represent these weights, which are defined in terms of the statistical properties of the images as

$$\mathbf{w}_{\text{IO},\beta} = \Sigma_{\beta}^{-1} \mathbf{s}. \quad (2)$$

The resulting Ideal observer strategy is implemented by comparing the weighted sum of the image, with mean background subtracted, to a detection threshold,

$$\begin{aligned} H_0 : & \text{if } \mathbf{w}_{\text{IO},\beta}^T \mathbf{g} < t_{\text{crit}} \\ H_1 : & \text{if } \mathbf{w}_{\text{IO},\beta}^T \mathbf{g} > t_{\text{crit}}. \end{aligned} \quad (3)$$

The value of the threshold,  $t_{\text{crit}}$ , determines the tradeoff between hits and false alarms. In principle this term should be set on the basis of outcome utilities. However, we leave it as a free parameter to be fit to the human observer data.

With human observer data, we obtain the equivalent contrast for the ideal observer by adjusting contrast and  $t_{\text{crit}}$  until the hit rate and false-alarm rate equal the human observer's. Let  $C_{\text{Obs},\beta}^D$  be the target contrast used for the human observer study, and let  $C_{\text{IO},\beta}^D$  be the equivalent ideal observer contrast. The efficiency of the observer is defined in terms of a squared ratio of contrast thresholds following Kersten (1987) as

$$\eta_{\text{obs},\beta}^D = \left( \frac{C_{\text{IO},\beta}^D}{C_{\text{Obs},\beta}^D} \right)^2. \quad (4)$$

Standard errors are determined by calculating efficiency on a session-by-session basis, and then computing the standard error across sessions.

### Localization tasks

As mentioned in the Introduction, the theory we use for ideal observers in a free-localization task comes from the work of Khurd and Gindi (2005). Here we present a somewhat simplified derivation that is adequate for our purposes. In this case, we have a conditional probability of the data for every possible location of the target. Let  $\mathbf{s}_l$  represent the profile of the target, when it is centered on the pixel with index  $l$ , which can be anywhere in

the search region (i.e.,  $128^2$  possible locations). The conditional likelihood of the data given a particular target location is

$$p(\mathbf{g}|l) = \text{MVN}(\mathbf{s}_l, \Sigma_{\beta}). \quad (5)$$

The basis for localization by the ideal observer is the posterior distribution on possible locations,  $p(l|\mathbf{g})$ . For a uniform prior distribution on target locations, the posterior distribution is proportional to the likelihood. Under the Gaussian assumptions of our images, we have

$$p(l|\mathbf{g}) = N_g e^{\mathbf{s}_l^T \Sigma_{\beta}^{-1} \mathbf{g}}, \quad (6)$$

where  $N_g$  is a normalization constant that ensures that  $p(l|\mathbf{g})$  sums to 1 over all possible locations.

The task specifies that any response within 5 pixels of the target center is considered a correct response. The ideal observer will therefore choose the location that maximizes the probability of a correct answer. For each point under consideration, the ideal observer adds up the probabilities of all points within a 5-pixel radius, to get a final score for the location. The point with the largest score is then chosen as the ideal observer's response. It is worth noting that the ideal observer response at a given location is very similar to an ideal detector with spatial uncertainty (Pelli, 1985), where uncertainty is confined to the acceptance region around a given location.

The ideal observer decision function can be implemented using convolutions to speed up the computationally intensive steps. For example, the stationary nature of the noise covariance matrix allows the computations of  $\mathbf{s}_l^T \Sigma_{\beta}^{-1} \mathbf{g}$  to be implemented by convolving the ideal observer template, defined in Equation 2, with the mean-subtracted background. Similarly the computation of the final score at each location in the image can be computed by convolving a disk of radius 5 pixels with the normalized posterior distribution in Equation 6. The recipe for computing the ideal observer begins with pre-computing the ideal observer filter by dividing the Fourier transform of the target by the power-spectrum of the noise. Then for each image, (1) this filter is used in a convolution after the mean background has been subtracted; (2) the result is exponentiated; (3) pixels outside the



search region are set to zero; (4) pixels in the search region are scaled so that they sum to 1; (5) the posterior is convolved with a disk of radius 5 pixels; and (6) the maximum point is chosen.

With a case-by-case ideal observer algorithm, the performance of the ideal observer is estimated to arbitrary accuracy using large sets of sample images. We use this approach to build LUTs of ideal observer performance as a function of target contrast. The LUTs for each  $\beta$  are determined in contrast increments of 0.01 from 0 until PC rises above 94%. The functions are plotted for each  $\beta$  in **Figure 4**. Each point is based on 2000 sample images, which results in standard errors that are less than 1% near the 80% correct level that is used in the experiments. Inverting these functions allows for us to determine the contrast threshold required by the ideal observer to achieve the specified level of PC. For an observer that achieves a proportion correct of  $PC_{Obs}$  in a localization task with a target contrast of  $C_{Obs,\beta}^L$ , efficiency is again defined (Kersten, 1987) as the squared contrast ratio

$$\eta_{Obs,\beta}^L = \left( \frac{C_{IO,\beta}^L(PC_{Obs})}{C_{Obs,\beta}^L} \right)^2. \quad (7)$$

Standard errors are determined by calculating efficiency on a session-by-session basis, and then computing the standard error across sessions.

The D&L task uses a similar process as the localization task, except that in the last step a threshold is applied. If the maximum score is above the detection threshold, the location of the score is

selected for detecting and localizing the target. If the maximum score is below the detection threshold, the ideal observer selects the “target-absent” response. For matching human observer data, the threshold contrast and detection criterion are adjusted to match the rate of correct detect-and-localize responses and the false positive (FP) rate. Efficiency is then calculated as the squared ratio of this contrast to the contrast used in the experiment, as in Equations 4 and 7.

## CLASSIFICATION IMAGES

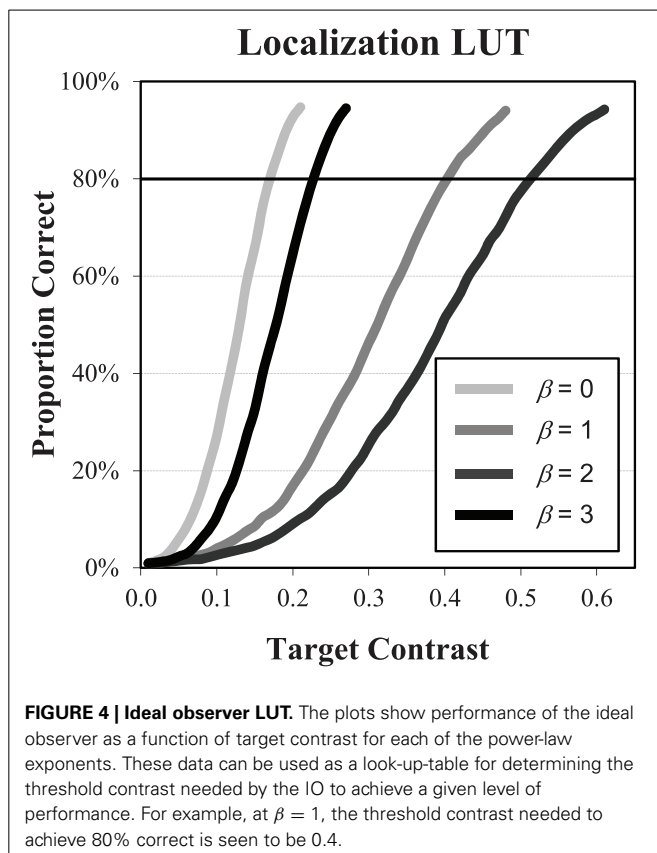
In addition to efficiency, we will use classification images as a way to investigate how visual processing affects task efficiency. This approach is straightforward for the detection tasks, where the classification image analysis has been well developed by Ahumada (2002) and others (Gold et al., 2000; Chauvin et al., 2005; Victor, 2005; Tjan and Nandy, 2006; Murray, 2011). Let  $\mathbf{n}$  represent the noise field for a given trial, with no target profile or mean background. Let us define the quantity  $\mathbf{q}$  as the product of the inverse covariance and the noise field,  $\mathbf{q} = \Sigma_{\beta}^{-1}\mathbf{n}$ . The classification image is given by

$$\mathbf{w}_{CI}^D = \bar{\mathbf{q}}_{FP} - \bar{\mathbf{q}}_{TN} + \bar{\mathbf{q}}_{TP} - \bar{\mathbf{q}}_{FN}, \quad (8)$$

where the  $\bar{\mathbf{q}}$  are the average  $\mathbf{q}$  over the FP, true-negative (TN), true-positive (TP), and false-negative (FN) noise fields. Under the (strong) assumption of a linear template as the mechanism for detecting the target, the classification image will provide an unbiased estimate of the template. If the observer does not follow the linear assumption, the resulting classification image may be distorted, depending on the degree of violation (Ahumada, 2002).

Tjan and Nandy (2006) have analyzed discrimination tasks in the presence of target location uncertainty using classification images. Their approach utilizes the concept of a “clamped signal,” in which the noise field masking the target profile in an incorrect response is analyzed. This approach was found to work well in various two-class detection and discrimination tasks with targets that could be subject to spatial uncertainty. Additionally, Neri (2010) has used early static nonlinearities as a way to model performance in such tasks. In principle, our free-localization task can be considered a classification task with  $128^2$  possible response categories (and a somewhat ambiguous definition of a correct response that includes neighboring locations). However, in this work we have pursued a different approach for classification images in which the noise at the location of an incorrect response is used rather than the noise that masked the unchosen target. In this regard, our approach is similar to a previous study by Rajashekar et al. (2006) that used eye-tracking to estimate gaze-contingent classification images, as well as studies that have used the classification-image approach in multiple-alternative forced choice studies (Caspi et al., 2004; Eckstein et al., 2007; Dai and Michéyl, 2010).

Let  $\mathbf{n}^A$  represent a “response-aligned” noise field, in which the image noise field is shifted so that the location selected by the observer is translated to the center of the image. Let  $\mathbf{q}^A = \Sigma_{\beta}^{-1}\mathbf{n}^A$ , which is analogous to a response-aligned version of  $\mathbf{q}$  defined above. For classification images in the localization tasks, we use



the average of the response-aligned  $\mathbf{q}$  vectors when the subject incorrectly localizes (IL) the target

$$\mathbf{w}_{\text{CI}}^{\text{I}} = \bar{\mathbf{q}}_{\text{FL}}^{\text{A}}. \quad (9)$$

In these cases, the response is entirely driven by the form of the noise at the response location. We will see below that this leads to a strong classification image relative to detection, even though the detection task uses all noise fields in the image and this approach for the free-localization uses approximately 20% of the trials in which a false-localization response is given.

As a simple test of the classification image approach for localization tasks, we have used it to evaluate the ideal observer. **Figure 5A** shows the frequency weights of the ideal observer, derived analytically from  $\Sigma_{\beta}^{-1}\mathbf{s}$ . In **Figure 5B**, we see the estimated frequency weights for 2000 trials of the ideal observer using Equation 9, when the target contrast is set so that PC = 80%. While there are some areas of apparent bias, particularly at the lowest spatial frequencies for  $\beta = 0$ , there is generally good agreement between the actual frequency weights used to perform the task and the estimated weights.

## RESULTS AND DISCUSSION

### PSYCHOMETRIC FUNCTIONS

Contrast thresholds were determined for each subject in each condition from fitted psychometric functions. After an initial training of 5 runs of increasing difficulty totaling 210 trials, psychometric data was acquired in 20 runs of 50 trials at five different contrast levels for a total of 200 trials at each contrast level. The contrast levels used were determined from pilot data. Cumulative Gaussian distribution functions were fit to the proportion of correct responses over the range of contrasts, and contrast thresholds were determined from the contrast that produced 80% correct. An example of the psychometric functions (Subject 4,  $\beta = 1$ ) is

shown in **Figure 6**. There was generally good agreement between the subject data and the cumulative Gaussian fitting function.

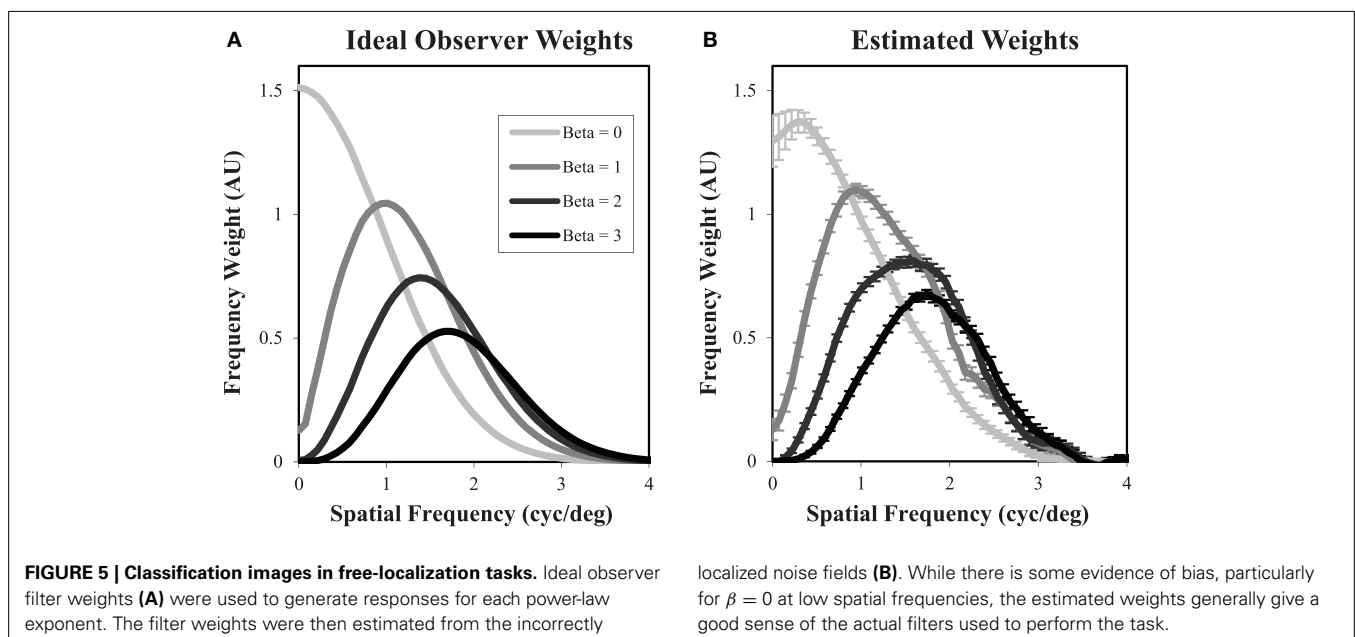
The average threshold contrast for each task and background type is plotted in **Figure 6B**. Thresholds within each task peak for  $b = 2$ . The thresholds are substantially higher for the localization task, with roughly a factor of two increase for each background.

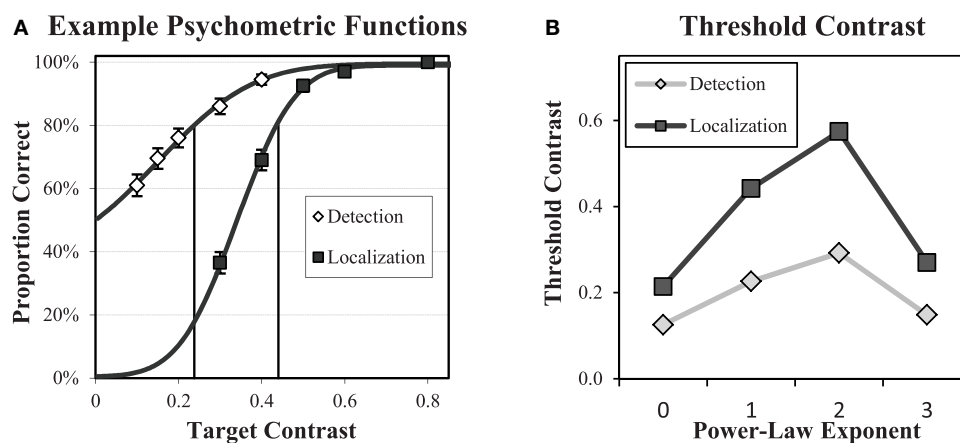
### CHARACTERIZING TASK PERFORMANCE

After each contrast threshold was determined from the psychometric data, subjects performed a total of 40 runs of 50 trials, for a total of 2000 trials at the subject's threshold contrast. Efficiency with respect to the ideal observer was estimated from this data. The efficiency results are described below in Detect-And-Localize Efficiency. Here, we will describe other measurements that provide additional information to characterize task performance.

Performance in the efficiency data is reasonably close to the nominal 80% correct levels derived from the psychometric functions. **Figure 7A** plots average PC across subjects from the efficiency data as a function of the power-law exponent of the background. Overall, PC values averaged 81.9% in the detection experiments and 80.3% in the localization experiments. The slight increases across subjects may be due to learning effects that occurred over the 2000 trials. The largest observed deviation from 80% correct for a single subject in a single condition was 7.3%. These results give us some confidence that efficiency was measured at contrasts near the actual 80% correct threshold.

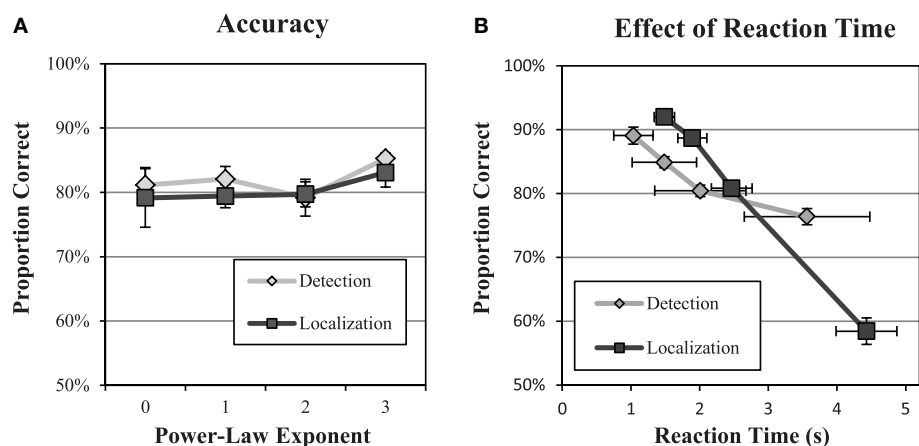
While reaction time is not an endpoint of our study, this data is recorded as part of the experimental procedure. Reaction time is defined as the time from stimulus onset to the acquisition of a subject response. Median reaction times, given in **Table 1**, are mostly larger for the free localization task. This is not surprising since the subject need to search an area  $6.7 \times 6.7^\circ$  in the localization task. Given the size of this area, the 48% average increase in reaction times seems rather modest. It is worth noting that the





**FIGURE 6 | Psychometric functions and thresholds.** An example of detection and forced-localization psychometric data (A) and fitted psychometric functions are shown for one subject in one condition. Error bars =  $\pm 1$  s.e. The fitting function is a cumulative Gaussian distribution that is used to determine the contrast threshold for 80% correct performance in the

subsequent experiments. The average subject contrast thresholds (B) in each power-law background is shown for both detection and localization tasks. Standard errors across subjects (not shown) are less than 0.01. The localization tasks requires approximately a factor of 2 greater contrast to obtain equivalent (80% correct) performance.



**FIGURE 7 | Accuracy and reaction time.** A check of performance levels in the efficiency data (A) shows that performance levels were reasonably close to the targeted 80% level. The midpoint

of reaction time in each quartile (B) is plotted against performance for the quartile. Averages and standard errors across subjects are shown.

**Table 1 | Reaction times.**

Subjects	Detection RT	Localization RT	Rel. dif. (%)
S1	1.35	2.61	94
S2	1.11	1.80	62
S3	1.17	1.77	51
S4	3.97	2.85	-28
S5	1.01	1.65	64
Ave	1.72	2.13	48

Median reaction times (RTs) are given for each subject as well as the relative difference between the detection and localization tasks.

increase in median response times is not uniform over the subjects. One subject (S4) is markedly slower in the detection task.

It is also of interest to compare the effect of reaction time and performance as shown in a representative example in **Figure 7B**.

We divided the data into quartiles of 500 trials according to reaction time, and then computed proportion correct in each quartile. The figure plots proportion correct as a function of the median reaction time for the quartile. All subjects exhibited a similar trend of decreased performance with greater reaction times in both tasks. This finding is the opposite of what might be expected from a speed-accuracy tradeoff, where slower speeds allow for more effective task performance. However, decreased performance for longer reaction times has been found previously (Eckstein et al., 2001), and is thought to reflect the effects of a noise limited task where longer reaction times are associated with noise masks that make the task more difficult.

Unlike the detection task, the localization response requires careful positioning of the cursor using the mouse. The accuracy of this process has consequences both for overall accuracy in the task, if mis-positioning the cursor causes the localization response

**Table 2 | Localization accuracy.**

Abs. dev.	$\beta = 0$	$\beta = 1$	$\beta = 2$	$\beta = 3$	Uniform
Pixels	$1.96 \pm 0.08$	$1.69 \pm 0.11$	$1.52 \pm 0.14$	$1.41 \pm 0.15$	3.40
Degrees	$0.102 \pm 0.004$	$0.088 \pm 0.006$	$0.079 \pm 0.007$	$0.074 \pm 0.008$	0.180

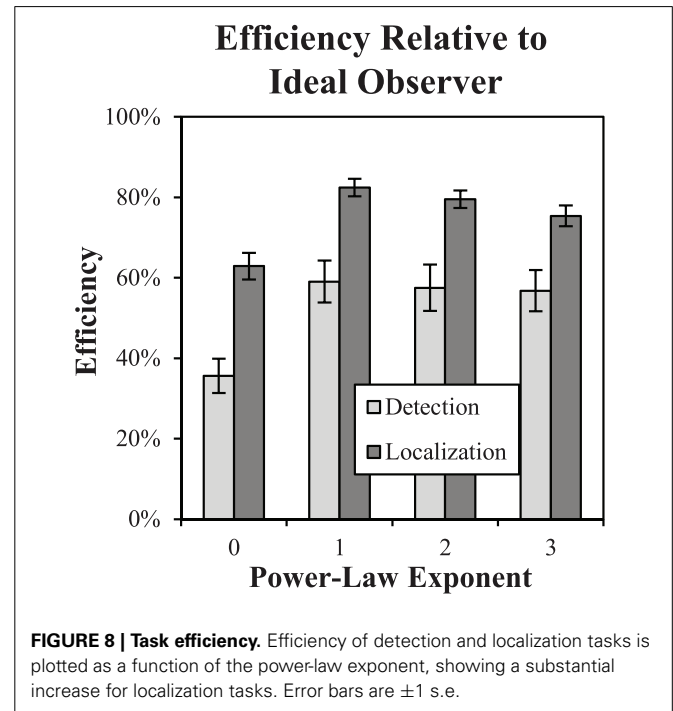
Average absolute deviation ( $\pm$ s.e. across subjects) of correct localization responses relative to the target center. Data is given both in pixel units as well as degrees of visual angle. For reference, the absolute deviation assuming a uniform distribution within the acceptance region is also given.

to fall outside the acceptance region, and for aligning the noise fields for the classification image analysis. To get some sense of the accuracy of the localization responses, we have evaluated the deviation of the responses, which is defined as the distance of the subject mouse clicks from the target location for responses that fall within the acceptance region of 5 pixels from the target center. **Table 2** gives the average deviation across subjects, in both pixels and degrees of visual angle, and well as the deviation assuming a uniform distribution of responses over the acceptance region. The deviations are all substantially smaller than the uniform distribution would predict, suggesting that there is considerable additional accuracy in the localization response. In addition, there is a consistent decrease in the deviation as  $\beta$  increases. The error represented by the absolute deviation contains both the effects of subject's misperception of the target center, as well as motor noise in the subject's response. Of these two, motor noise will be detrimental to the classification image methodology, since it will lead to misalignment of the selected noise fields. The observed deviations in **Table 2** act as an upper bound on motor noise in the subject responses, and suggest that these effects may be modest.

### TASK EFFICIENCY

The primary performance result we are interested in for these studies is observer efficiency, as plotted in **Figure 8**. Efficiency with respect to the ideal observer appears to be substantially higher for localization tasks than detection tasks. A Two-Way ANOVA with the five subjects considered as replications finds significant effects for both the task [ $F_{(1,32)} = 63.4$ ,  $p < 0.0001$ ] as well as the background exponent [ $F_{(3, 32)} = 11.7$ ,  $p < 0.0001$ ]. The interaction between task and exponent was not found to be significant [ $F_{(3, 32)} = 0.39$ ,  $p > 0.76$ ]. It should be noted that average efficiency near 80% for  $\beta$ -values of 1–3, is considered quite high. In the classic experiments by Burgess et al. (1981), efficiency as high as 70% was observed with averages across observers closer to 50%. These experiments used a spatial forced choice methodology and white noise ( $\beta = 0$ ). Experiments in low-pass noise similar to the  $\beta = 3$  condition used here (Abbey and Eckstein, 2007), found efficiency in the 40 to 60% range. These are consistent with our findings in the detection task, all of which utilize aperiodic “bump” targets. Efficiency of oscillatory targets are typically lower (Legge et al., 1987). The increased efficiency we find in the localization tasks represents a substantial gain from these fixed-location tasks, and suggests that subjects have little room for sub-optimal computations in performing these tasks.

Efficiency is somewhat lower for  $\beta = 0$  in both the detection and localization tasks. We consider this case further in Efficiency of Image Processing for  $\beta = 0$ , below. We also note that these



efficiency values appear to be relatively stable with the acceptance radius. We observed less than a 1% difference in observed efficiency varying the acceptance region from 4 pixels to 7.

These efficiency results show that in spite of larger thresholds for the free-localization relative to detection, as shown in **Figure 6B**, overall efficiency is substantially higher. This means that thresholds for the ideal observer increase proportionally even more than the human subjects' did. Our findings are consistent with the uncertainty hypothesis (Tanner, 1961; Pelli, 1985), which posits imperfect use of the location cues in detection tasks, and leave the observer with some residual uncertainty regarding the location of the target that can reduce performance. The ideal observer is not subject to this phenomenon, which results in a somewhat lower contrast threshold. In the free-localization task, where uncertainty is intrinsic to the task, the ideal observer does not have the advantage of precise knowledge of location, and contrast thresholds rise relative to the human observers as a result. However, other explanations for the large difference in efficiency are possible. For example, detection tasks require that the subject use some sort of criterion that dichotomizes responses. If this criterion drifts or is prone to jitter, performance will be reduced. This possibility motivated the detect-and-localize study.

### DETECT-AND-LOCALIZE EFFICIENCY

A subset of three subjects performed the detect-and-localize experiments, which were all run after the detection and localization data were acquired. Threshold target contrast from the localization tasks were used as target contrasts for these experiments. The proportion of correct responses dropped modestly from an average of 80.3% in the localization tasks to 76% in the D&L tasks. **Figure 9** plots shows the efficiency data for the detection task, localization task, and D&L tasks as a function of the power-law exponent for the subset of subjects that participated in all three studies. The average efficiency values for the D&L are all well above both the detection and localization tasks. In fact several observed values are near 90% efficiency, which is again quite high for tasks masked by luminance noise. These findings are close to the highest reported efficiency we are aware of for visual tasks limited by noise (Manjeshwar and Wilson, 2001).

### EFFICIENCY OF IMAGE PROCESSING FOR $\beta = 0$

**Figure 8** shows reduced efficiency in the  $\beta = 0$  condition of both the detection and localization tasks. After finding this effect, we were interested in whether it might be mitigated by processing the images to have a background power-law of  $\beta = 2$ , where efficiency was generally better. As described above in Stimulus and Display Properties, this is accomplished by filtering the images with a kernel that has a  $1/f$  spectrum, which will modify both the background statistics and the target profile, as shown in **Figure 3**.

**Figure 10** shows that the effect of processing is to bring efficiency in the  $\beta = 0$  condition up to 66% in the detection task

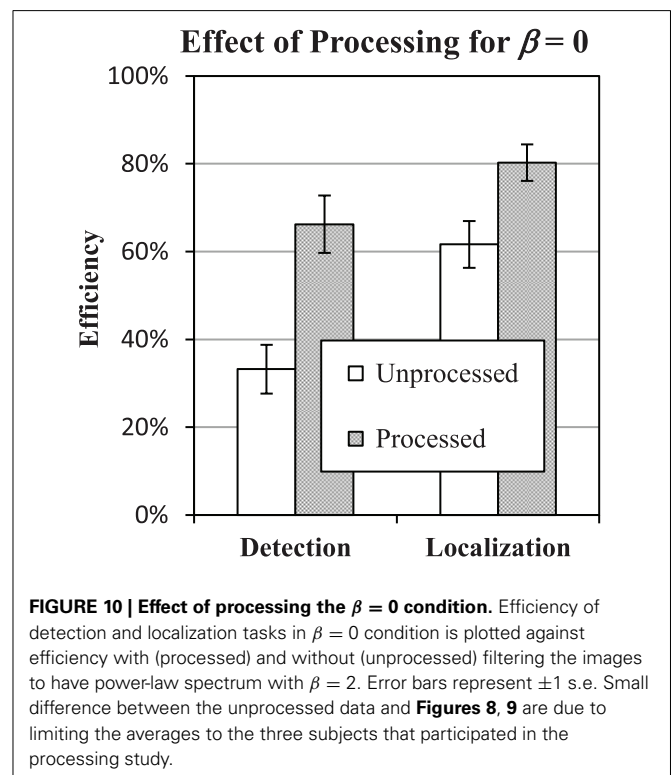
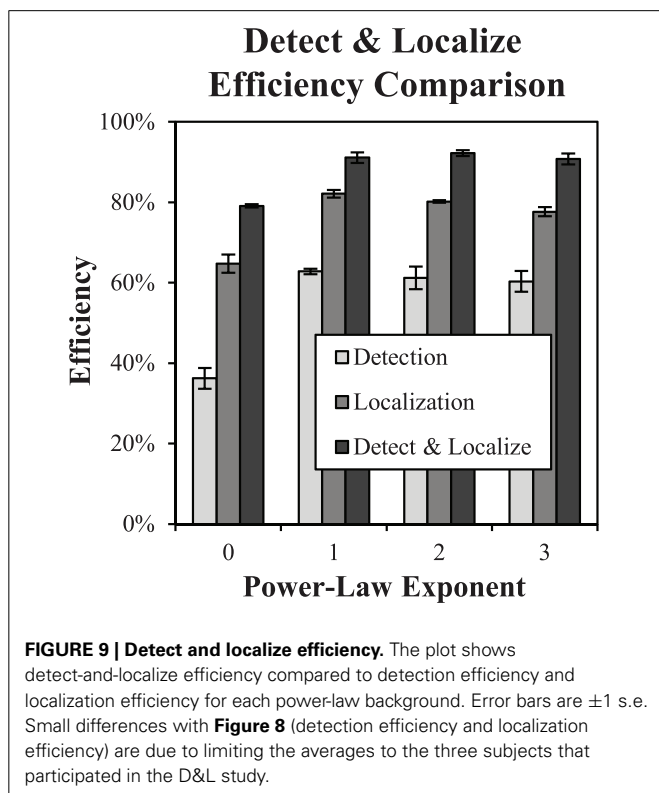
and 80% in the localization task. These levels are consistent with efficiency levels found for  $\beta$  in the range of 1–3.

### CLASSIFICATION IMAGES

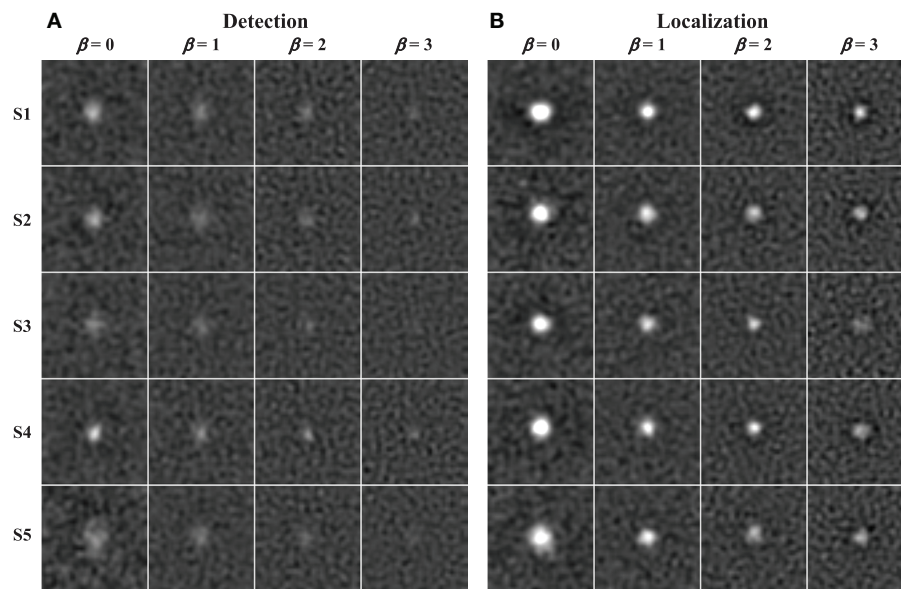
**Figure 11** shows the classification images for each subject in each background condition for both the detection and localization tasks. The images are cropped to the central  $2.1^\circ$  of visual angle (40 pixels). Outside of this area, there are no discernable features beyond what appears to be estimation error in the classification images. To mitigate the effects of noise, the classification images have been low-pass filtered with a 4th-order Butterworth filter, with the roll-off parameter set to 5.6 cyc/deg (0.29 cyc/pixel). This was well beyond the point at which the spatial frequency plots below appear to decay to zero.

The images in **Figure 11** were windowed to have approximately the same mean background and error magnitude. Thus, the intensity of the features in the observed classification images gives some sense of their signal-to-noise ratio (SNR). The generally brighter appearance of classification images in the localization tasks relative to the corresponding detection tasks suggests that search process may lead to methodological advantages for estimating classification images, even though the localization classification images are estimated from approximately 20% of the subjects responses in which an incorrect localization response is given. There also appears to be some differences in the intensity of the classification images going from  $\beta = 0$  to  $\beta = 3$ , and there are clearly individual differences between subjects.

In addition to the overall intensity of the classification images, we are also interested in the profile of these decision weights.







**FIGURE 11 | Classification images.** Estimated classification images (cropped to  $2.1^\circ$  per side) are shown for each condition (columns) and subject (rows) in the detection (A) and localization (B) tasks. The images are windowed to have approximately the same magnitude of estimation error.

Based on previous experience, we find that differences between classification images in different conditions are most clearly depicted for radial averages in the spatial-frequency domain. **Figure 12** plots the classification frequency weights averaged over subjects and normalized so that the weight at the peak frequency is 1. To reduce the effects of noise in the classification images, a Butterworth spatial window with a cutoff of  $1.05^\circ$  (20 pixels) was applied before the Fourier transform and radial averaging. For reference, we have plotted the classification weights of the ideal observer as well. In all conditions, the average frequency weights assume a bandpass form, peaking at frequencies between 0.7 cyc/deg and 1.6 cyc/deg as  $\beta$  goes from 0 to 3. As has been found previously (Abbey and Eckstein, 2007; Conrey and Gold, 2009), the classification weights here give evidence of visual processing that is changing with the different power-law textures in the background. But this process is not as extreme as the adaptation that occurs in the ideal observer, where peak frequencies move from 0 to 1.7 cyc/deg.

In the  $\beta = 0$  condition, we observe substantial underweighting of low spatial frequencies relative to the ideal observer. Of interest for the comparison of detection and localization tasks, there is less low-frequency suppression in the localization task compared to the detection task. As  $\beta$  increases, we see that the low-frequency profiles come together, but now they do not suppress low frequencies as much as the ideal observer. Also, as  $\beta$  increases, the classification weight frequency profiles begin to diverge at higher spatial frequencies above the peak values. Here the profiles from the localization tasks have higher weights that are closer to the ideal observer.

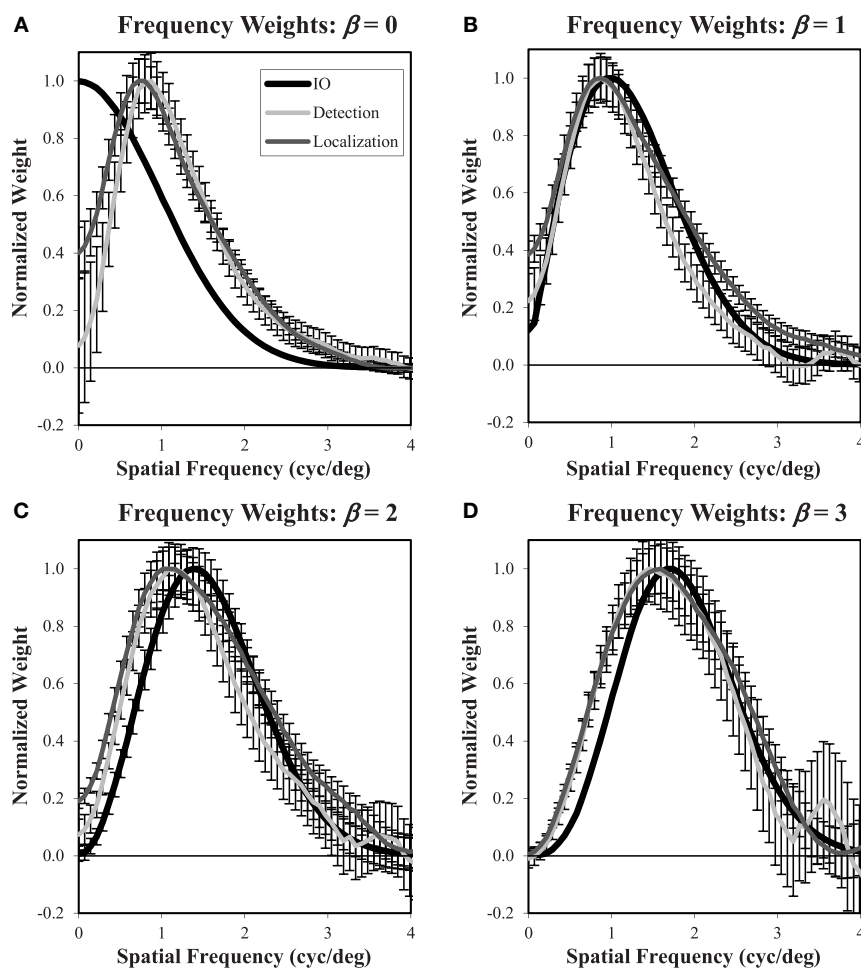
**Figure 13** shows the frequency plots in the  $\beta = 0$  condition using responses from the processed and unprocessed data averaged over the three subjects that participated in these studies.

The plots show processing effectively modifies the weighting profile that subjects use. In both tasks, the effect of processing is to increase the low-frequency weighting so that the average subject classification weights more closely match the ideal observer. Thus, the classification image profiles give a visual mechanism for the improved efficiency found in **Figure 10**.

## SUMMARY AND CONCLUSIONS

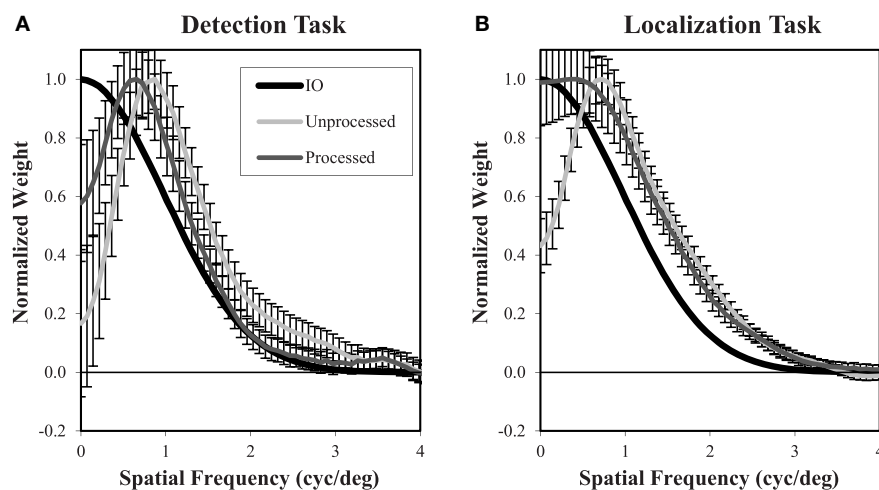
We find human observers substantially improve in performance relative to the ideal observer in free-localization tasks compared to fixed-location detection tasks, in spite of increased contrast thresholds. This occurs in all four power-law textures that were investigated. In a follow-up study investigating a detect-and-localize task, we find the highest measured efficiency in our experiments, suggesting that our efficiency results are not simply a consequence of a general inability to maintain detection criteria. Our findings are consistent with spatial uncertainty as a limiting effect in the presence of location cues.

While it is clear from the classification images that observers are able to tune their visual templates to the statistics of the noise in the images, there is also evidence that this process is limited in both fixed and free-localization tasks. Despite a common target profile, the different power-law textures require different frequency tuning to achieve optimal performance. We do find some evidence of such tuning in the classification images estimated from the subject responses. Peak spatial frequency weights change by roughly a factor of two going from a power-law exponent of  $\beta = 0$  to  $\beta = 3$  (0.72–1.59 cyc/deg). However, on average the subject frequency weights exhibited some clear departures from optimal tuning as defined by the ideal observer. At  $\beta = 0$ , we find human observer frequency weights shifted to higher spatial frequencies relative to the ideal observer. For  $\beta > 0$ ,



**FIGURE 12 | Frequency weights derived from Classification images.** Radial frequency profiles are shown for each of the four power-law textures (A–D) with normalization so that the maximum weight is one.

The ideal observer profile is derived from theory. The detection and localization plots are averaged across the five subjects. Error bars are  $\pm 1$  s.e. averaged across subjects. The legend (A) applies to all plots.



**FIGURE 13 | Frequency weights for processed and unprocessed images.** These plots are similar to Figure 12 and show estimated weights from the  $\beta = 0$  images using the responses to processed and unprocessed images. In

both the detection (A) and localization (B) tasks, the effect of image processing is to increase the estimated weights at low spatial frequency, bringing them closer to the ideal observer weights. The legend (A) applies to both plots.

human-observer classification weights peak at lower spatial frequencies than the ideal observer.

Frequency tuning of subjects in the white-noise condition was most different from the ideal observer. This condition also led to the lowest efficiency in performance. Since  $\beta = 0$  was the power-law exponent furthest from that found in natural scenes ( $\beta = 2$ ), this finding is consistent with the idea that the human visual system is somewhat adapted to the statistics of natural images. The follow-up study investigating processed images supports this connection by finding uniformly improved performance when the white-noise images were filtered to have  $\beta = 2$ . Filtering the images was also seen to effectively improve frequency tuning of the subjects in the white-noise condition.

While we do not attempt to explicitly model the visual system to explain our findings, we do believe that our findings may be relevant in such attempts, for the same reasons given originally by Burgess (Burgess et al., 1981). The finding of high efficiency in free-localization and detect-and-localize tasks suggest that models of vision in these tasks cannot be very different, at a computational level, from the ideal observer, and thus may provide a valuable constraint to such efforts in future studies.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the U.S. National Institutes of Health (NEI-EY015925).

## REFERENCES

- Abbey, C. K., and Eckstein, M. P. (2007). Classification images for simple detection and discrimination tasks in correlated noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 24, B110–B124. doi: 10.1364/JOSAA.24.00B110
- Abbey, C. K., Zemp, R. J., Liu, J., Lindfors, K. K., and Insana, M. F. (2006). Observer efficiency in discrimination tasks simulating malignant and benign breast lesions imaged with ultrasound. *IEEE Trans. Med. Imaging* 25, 198–209. doi: 10.1109/TMI.2005.862205
- Ahumada, A. J. Jr. (2002). Classification image weights and internal noise level estimation. *J. Vis.* 2, 121–131. doi: 10.1167/2.1.8
- Barlow, H. (1977). Retinal and central factors in human vision limited by noise. *Vertebrate Photoreception* 337, C358.
- Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Res.* 18, 637–650. doi: 10.1016/0042-6989(78)90143-8
- Barlow, H. B., and Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Res.* 19, 783–793. doi: 10.1016/0042-6989(79)90154-8
- Bochud, F. O., Abbey, C. K., and Eckstein, M. P. (2004). Search for lesions in mammograms: statistical characterization of observer responses. *Med. Phys.* 31, 24–36. doi: 10.1118/1.1630493
- Bochud, F. O., Valley, J. F., Verdun, F. R., Hessler, C., and Schnyder, P. (1999). Estimation of the noisy component of anatomical backgrounds. *Med. Phys.* 26, 1365–1370. doi: 10.1118/1.598632
- Burgess, A. E., and Ghandeharian, H. (1984). Visual signal detection. II. Signal-location identification. *J. Opt. Soc. Am. A* 1, 906–910. doi: 10.1364/JOSAA.1.000906
- Burgess, A. E., Jacobson, F. L., and Judy, P. F. (2001). Human observer detection experiments with mammograms and power-law noise. *Med. Phys.* 28, 419–437. doi: 10.1118/1.1355308
- Burgess, A. E., Wagner, R. F., Jennings, R. J., and Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science* 214, 93–94. doi: 10.1126/science.7280685
- Burton, G. J., and Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Appl. Opt.* 26, 157–170. doi: 10.1364/AO.26.000157
- Caspi, A., Beutter, B. R., Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13086–13090. doi: 10.1073/pnas.0305329101
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M., and Gosselin, F. (2005). Accurate statistical tests for smooth classification images. *J. Vis.* 5, 659–667. doi: 10.1167/5.9.1
- Chen, L., Abbey, C. K., and Boone, J. M. (2013). Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities. *Phys. Med. Biol.* 58, 1663–1681. doi: 10.1088/0031-9155/58/6/1663
- Chen, L., Abbey, C. K., Nosrateih, A., Lindfors, K. K., and Boone, J. M. (2012). Anatomical complexity in breast parenchyma and its implications for optimal breast imaging strategies. *Med. Phys.* 39, 1435–1441. doi: 10.1118/1.3685462
- Conrey, B., and Gold, J. M. (2009). Pattern recognition in correlated and uncorrelated noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 26, B94–B109. doi: 10.1364/JOSAA.26.000B94
- Dai, H., and Michéyl, C. (2010). Psychophysical reverse correlation with multiple response alternatives. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 976–993. doi: 10.1037/a0017171
- Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaiki, S. S., and Stone, L. S. (2007). Similar neural representations of the target for saccades and perception during search. *J. Neurosci.* 27, 1266–1270. doi: 10.1523/JNEUROSCI.3975-06.2007
- Eckstein, M. P., Beutter, B. R., and Stone, L. S. (2001). Quantifying the performance limits of human saccadic targeting during visual search. *Perception* 30, 1389–1401. doi: 10.1068/p3128
- Engstrom, E., Reiser, I., and Nishikawa, R. (2009). Comparison of power spectra for tomosynthesis projections and reconstructed images. *Med. Phys.* 36, 1753–1758. doi: 10.1118/1.3116774
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394. doi: 10.1364/JOSAA.4.002379
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314. doi: 10.1037/0033-295X.96.2.267
- Gold, J. M., Murray, R. F., Bennett, P. J., and Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Curr. Biol.* 10, 663–666. doi: 10.1016/S0960-9822(00)00523-6
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.
- Insana, M. F., and Hall, T. J. (1994). Visual detection efficiency in ultrasonic imaging: a framework for objective assessment of image quality. *J. Acoust. Soc. Am.* 95, 2081. doi: 10.1121/1.408669
- Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Res.* 27, 1029–1040. doi: 10.1016/0042-6989(87)90016-2
- Khurd, P., and Gindi, G. (2005). Decision strategies that maximize the area under the LROC curve. *IEEE Trans. Med. Imaging* 24, 1626–1636. doi: 10.1109/TMI.2005.859210
- Khurd, P., Liu, B., and Gindi, G. (2010). Ideal AFROC and FROC observers. *IEEE Trans. Med. Imaging* 29, 375–386. doi: 10.1109/TMI.2009.2031840
- Legge, G. E., Kersten, D., and Burgess, A. E. (1987). Contrast discrimination in noise. *J. Opt. Soc. Am. A* 4, 391–404. doi: 10.1364/JOSAA.4.000391
- Liu, B., Zhou, L., Kulkarni, S., and Gindi, G. (2009). The efficiency of the human observer for lesion detection and localization in emission tomography. *Phys. Med. Biol.* 54, 2651–2666. doi: 10.1088/0031-9155/54/9/004
- Manjeshwar, R. M., and Wilson, D. L. (2001). Hyperefficient detection of targets in noisy images. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 18, 507–513. doi: 10.1364/JOSAA.18.000507
- Metheany, K. G., Abbey, C. K., Packard, N., and Boone, J. M. (2008). Characterizing anatomical variability in breast CT images. *Med. Phys.* 35, 4685–4694. doi: 10.1118/1.2977772
- Murray, R. F. (2011). Classification images: a review. *J. Vis.* 11, 1–25. doi: 10.1167/11.5.2
- Myers, K. J., Barrett, H. H., Borgstrom, M. C., Patton, D. D., and Seeley, G. W. (1985). Effect of noise correlation on detectability of disk signals in medical imaging. *J. Opt. Soc. Am. A* 2, 1752–1759. doi: 10.1364/JOSAA.2.001752
- Neri, P. (2010). Visual detection under uncertainty operates via an early static, not late dynamic, non-linearity. *Front. Comput. Neurosci.* 4:151. doi: 10.3389/fncom.2010.00151
- Park, S., Clarkson, E., Kupinski, M. A., and Barrett, H. H. (2005). Efficiency of the human observer detecting random signals in random backgrounds. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 22, 3–16. doi: 10.1364/JOSAA.22.000003

- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* 2, 1508–1532. doi: 10.1364/JOSAA.2.001508
- Rajashekar, U., Bovik, A. C., and Cormack, L. K. (2006). Visual search in noise: revealing the influence of structural cues by gaze-contingent classification image analysis. *J. Vis.* 6, 379–386. doi: 10.1167/6.4.7
- Siewerdsen, J. H., and Jaffray, D. A. (2000). Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography). *Med. Phys.* 27, 1903–1914. doi: 10.1118/1.1286590
- Tanner, W. P. Jr. (1961). Physiological implications of psychophysical data. *Ann. N.Y. Acad. Sci.* 89, 752–765. doi: 10.1111/j.1749-6632.1961.tb20176.x
- Tjan, B. S., Braje, W. L., Legge, G. E., and Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Res.* 35, 3053–3069. doi: 10.1016/0042-6989(95)00070-G
- Tjan, B. S., and Nandy, A. S. (2006). Classification images with uncertainty. *J. Vis.* 6, 387–413. doi: 10.1167/6.4.8
- Victor, J. D. (2005). Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy. *Nat. Neurosci.* 8, 1651–1656. doi: 10.1038/nn1607
- Wagner, R. F., and Brown, G. G. (1985). Unified SNR analysis of medical imaging systems. *Phys. Med. Biol.* 30, 489–518. doi: 10.1088/0031-9155/30/6/001
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 27 January 2014; paper pending published: 06 March 2014; accepted: 02 April 2014; published online: 01 May 2014.
- Citation: Abbey CK and Eckstein MP (2014) Observer efficiency in free-localization tasks with correlated noise. *Front. Psychol.* 5:345. doi: 10.3389/fpsyg.2014.00345
- This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.
- Copyright © 2014 Abbey and Eckstein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Motion processing: the most sensitive detectors differ in temporally localized and extended noise

Rémy Allard<sup>1,2,3 \*</sup> and Jocelyn Faubert<sup>4,5</sup>

<sup>1</sup> INSERM, U968, Paris, France

<sup>2</sup> Institut de la Vision, UMR\_S 968, Sorbonne Universités – Université Pierre-et-Marie-Curie, Paris, France

<sup>3</sup> CNRS, UMR\_7210, Paris, France

<sup>4</sup> Visual Psychophysics and Perception Laboratory, Université de Montréal, Montréal, QC, Canada

<sup>5</sup> NSERC-Essilor Industrial Research Chair, Montreal, QC, Canada

## Edited by:

Denis Pelli, New York University, USA

## Reviewed by:

Richard J. A. Van Wezel, Radboud University, Netherlands  
Szonya Durant, Royal Holloway, University of London, UK

## \*Correspondence:

Rémy Allard, Institut de la Vision, UMR\_S 968, Sorbonne Universités – Université Pierre-et-Marie-Curie, 17 rue Moreau, 75012 Paris, France  
e-mail: remy.allard@inserm.fr;  
Website: www.faubertlab.com

Contrast thresholds for discriminating orientation and direction of a drifting, oriented grating are usually similar to contrast detection thresholds, which suggest that the most sensitive detectors are labeled for both orientation and direction (Watson and Robson, 1981). This was found to be true in noiseless condition, but Arena et al. (2013) recently found that this was not true in localized noise (i.e., noise having the same spatiotemporal window as the target) as thresholds for discriminating direction were higher than for discriminating orientation. They suggested that this could be explained by the fact that there are more neurons selective to orientation than direction. Another possible interpretation is that, unlike contrast thresholds in absence of noise, the most sensitive detectors in localized noise were labeled for orientation, but not for direction. This hypothesis is supported by recent findings showing different processes operating in localized and extended noise (i.e., full-screen, continuously displayed noise, Allard and Cavanagh, 2011). In the current study, we evaluated contrast thresholds for orientation and direction discrimination tasks in noiseless conditions, and in noise that was either spatially localized or extended, and temporally localized or extended. We found similar orientation and direction thresholds in absence of noise and in temporally extended noise, but greater direction thresholds in temporally localized noise. This suggests that in noiseless and temporally extended noise the most sensitive detectors were labeled for both orientation and direction (e.g., direction-selective complex cells), whereas in temporally localized noise the most sensitive detectors were labeled for orientation but not direction (e.g., simple cells). We conclude that to avoid violating the noise-invariant processing assumption, external noise paradigms investigating motion processing should use noise that is temporally extended, not localized.

**Keywords:** local noise, extended noise, motion, detection, discrimination

## INTRODUCTION

Arena et al. (2013) used an external noise paradigm to investigate age-related sensitivity losses to motion processing by measuring contrast thresholds for discriminating either the orientation or the direction of drifting gratings. When the dominating noise source was internal because external noise had a negligible impact (i.e., in low noise), they observed an age-related sensitivity loss for both tasks, which could be due, according to the linear amplifier model (Pelli, 1981; Pelli and Farell, 1999), to an increase in internal equivalent noise (e.g., more internal noise with aging) or a decrease in calculation efficiency (i.e., greater signal-to-noise ratios required to perform the tasks with aging). Conversely, when internal noise had a negligible impact because the dominating noise source was external (i.e., in high noise), the two age groups had similar contrast thresholds and thereby necessitated similar signal-to-noise ratios to perform the tasks (i.e., they had similar calculation efficiencies). By implicitly assuming that the calculation efficiency in low noise was the same as the measured calculation efficiency in high noise (i.e., the noise-invariant processing assumption underlying external noise paradigms, Allard and Cavanagh, 2011), they

concluded that the age-related sensitivity losses in low noise were due to an increase in internal equivalent noise, not a decrease in calculation efficiency.

However, their data actually suggest that different processes operated in low and high noise, which would invalidate the assumption that the calculation efficiencies in low noise were the same as the measured calculation efficiencies in high noise. In low noise, similar contrast thresholds were observed for discriminating orientation and direction, which suggests that both measured the sensitivity of the same processing system having its most sensitive detectors labeled for both orientation and direction (Watson and Robson, 1981). In high noise, however, contrast thresholds for orientation discrimination were lower than for direction discrimination suggesting that the most sensitive detectors were labeled for orientation, but not direction. Consequently, in low noise the most sensitive detectors would be labeled for both orientation and direction (e.g., direction-selective complex cells), but in high noise they would only be labeled for orientation, not for direction (e.g., simple cells). If contrast thresholds depended on the sensitivity of different detectors in low and



high noise, then the assumption that the calculation efficiency in low noise was the same as the measured calculation efficiency in high noise would be compromised and without knowing the calculation efficiency in low noise it is not possible to determine if the age-related sensitivity loss in low noise was due to an increase in internal equivalent noise or a decrease in calculation efficiency.

As in many studies, Arena et al. (2013) used spatiotemporally localized noise appearing only at the spatiotemporal target location (personal communication), which could explain that the most sensitive detectors were not the same in low and high noise. Indeed, given that internal noise (which dominates in low noise) is spatiotemporally extended (e.g., it does not turn on and off with the stimulus and it is not located only at the stimulus location), the dominating noise source in low and high localized noise have different spatiotemporal windows: extended in low noise and localized in high noise. If the most sensitive detectors differ depending on whether the dominating noise source is localized or extended, this would cause the most sensitive detectors to differ in low and high localized noise, which would compromise the assumption that the calculation efficiency in low noise (i.e., in extended internal noise) is the same as the measured calculation efficiency in high localized noise. For instance, noise that is temporally localized to the target (i.e., turn on and off with the target) introduces strong onset and offset transients, which could result in a greater masking effect on direction selective detectors making detectors labeled for orientation more sensitive than detectors labeled for both orientation and direction.

The objective of the present study was to determine if the processes (e.g., most sensitive detectors) involved in discriminating the orientation and the direction of drifting gratings in localized and extended noise differ from the processes operating in absence of noise. More specifically, the goal of the current study was to determine whether the calculation efficiencies in absence of noise (i.e., in extended internal noise) differ for orientation and direction discrimination (as observed by Arena et al. in high localized noise) or not (as suggested by the similar contrast thresholds in low noise). To investigate this, we conducted an experiment similar to Arena et al.'s (2013) in which contrast thresholds were measured for discriminating orientation and direction in absence of noise (i.e., in extended internal noise) and in high noise having different spatiotemporal windows: spatially localized or extended and temporally localized or extended. Given that contrast threshold depends on both the dominating noise source and the calculation efficiency (i.e., signal-to-noise ratio required to perform the task) and that the level of the dominating noise source is known in high noise, calculation efficiency in high noise can be directly measured by measuring contrast threshold in high noise. If the calculation efficiencies in absence of noise (which cannot be directly measured because the internal noise level is unknown) are the same for orientation and direction discrimination, but differ in high localized noise (as measured by Arena et al., 2013) due to the noise being localized, then we would expect the calculation efficiencies in high extended noise to be similar for orientation and direction discrimination. This would show a violation of the noise-invariant processing assumption when using

localized noise as the calculation efficiencies measured in localized noise would not reflect the calculation efficiencies in absence of noise. Conversely, if the calculation efficiency in absence of noise is greater for orientation discrimination (as measured in high localized noise by Arena et al., 2013), then the calculation efficiency for orientation discrimination should also be greater in high extended noise. For instance, Arena et al. (2013) hypothesized that the calculation efficiency difference between the two tasks in high localized noise could be due to more neurons responding to orientation than to direction or to the fact that discriminating direction requires more spatiotemporal integration than discriminating orientation. In either case, a similar calculation efficiency (i.e., contrast threshold) difference would also be expected in extended noise.

The current study is not the first to question the use of localized noise within external noise paradigms. Allard and Cavanagh (2011) found that crowding impaired contrast detection in the near periphery in localized noise, but not in absence of noise or in extended noise. We found that aging can impair contrast thresholds in localized noise, but not in extended noise (Allard et al., 2013). Furthermore, we recently argued that using spatiotemporally localized noise that is also localized as a function of orientation and frequency (i.e., contains only the orientation and frequency of the stimulus) makes a contrast detection task switch to a contrast discrimination task (Allard and Faubert, 2013). All these studies focused on the contrast detection of a static target. However, because localized noise introduces strong transients, using localized noise could be even more critical for motion processing. The current study addresses this question using a different paradigm that more directly identifies the underlying process (compared to crowding or aging) by determining if the most sensitive detectors are direction selective or not.

## MATERIALS AND METHODS

### OBSERVERS

Three naïve observers, who were financially compensated and provided informed consent, and one of the authors, participated in this study. They had normal or corrected-to-normal vision.

### APPARATUS

The stimuli were presented on a 19-inch CRT monitor with a refresh rate of 120 Hz. The Noisy-Bit method (Allard and Faubert, 2008) implemented independently to each gun made the 8-bit display perceptually equivalent to an analog display having a continuous luminance resolution. The monitor was the only source of light in the room. A Minolta CS100 photometer interfaced with a homemade program calibrated the output intensity of each gun. At the viewing distance of 114 cm, the width and height of each pixel were  $1/64^\circ$  of visual angle.

### STIMULI AND PROCEDURE

The signal was a 0.5 cpd sine wave grating drifting at a frequency of 1.875 Hz. Observers were asked to report either the orientation (tilted either  $-45^\circ$  or  $45^\circ$  from vertical) or the drifting direction. When the task was to report the orientation, both the orientation ( $-45^\circ$  or  $45^\circ$ ) and direction were randomized. When the task was

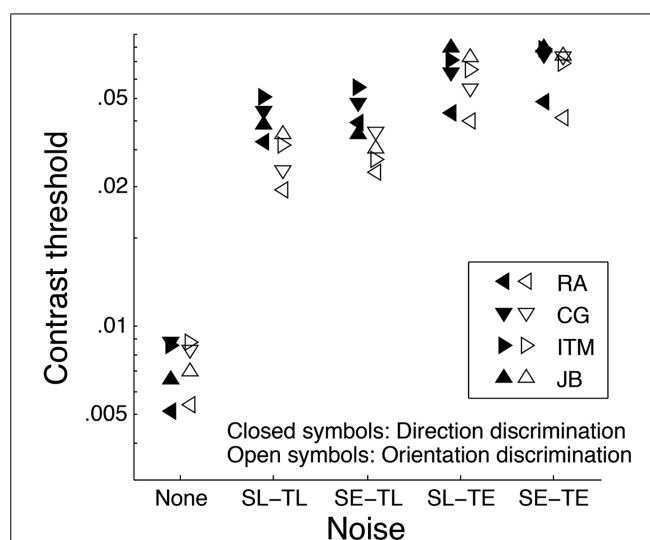
to report the drifting direction, the orientation was fixed for a given block of trials and the drifting direction was randomized. The initial phase of the grating was randomized on each trial. The signal was presented for 267 ms. The spatial window was a circular aperture of  $4^\circ$  plus a half-cosine edge of  $0.5^\circ$ . The contrast was controlled by a 3-down-1-up staircase procedure (Levitt, 1971) with step size of 0.1 log, which was interrupted after 100 trials. The contrast threshold for a given staircase was estimated as the geometric mean of the inversions.

There were five different noise conditions: no noise and four noise conditions resulting from the combinations of two spatial and two temporal windows. The spatial window was either localized or extended, i.e., the same as the signal window or full-screen, respectively. The temporal window was also either localized or extended, i.e., turn on and off with the signal or continuously present (including between trials), respectively. The noise was binary with element size of  $4 \times 4$  pixels (i.e.,  $0.0625 \times 0.0625^\circ$ ) and resampled every other frame (i.e., dynamic at 60 Hz). Thus, the fact that the noise was not correlated over space (across noise elements) and time (across frames) implies that it was both temporally and spatially white, that is, it had the same spectral energy at all frequencies (within the limit of the spatial and temporal resolution of the noise). The noise was superimposed to the signal (both summed) and to avoid luminance motion drifting cues within noise elements, there was no spatial or temporal luminance variation within each noise element.

For each noise condition, contrast thresholds were estimated for direction and orientation discrimination. To perform the same number of measurements for orientation and direction discrimination, a given noise block contained four staircases: direction discrimination for the two orientations ( $-45^\circ$  and  $45^\circ$ ) and two identical orientation discriminations. The four staircases were blocked and tested in a random order. Each of the five noise blocks was tested twice in a pseudo-random order resulting in 10 noise blocks (two blocks per noise condition) each composed of four staircases (two for direction discrimination and two for orientation discrimination) performed in a random order (not interlaced). As a result, for each noise condition, the two contrast threshold estimations were based on the geometric mean of the contrast thresholds estimations based on four staircases.

## RESULTS

**Figure 1** shows contrast thresholds for orientation (open symbols) and direction (filled symbols) discrimination. Contrast thresholds in the four conditions with noise were substantially higher (by a factor of about 4) than the condition without noise. This confirms that these four conditions were performed in high noise, that is, the impact of internal noise was negligible (i.e., the dominating noise source was external) so that contrast thresholds therefore depended solely on calculation efficiency, not on internal equivalent noise. Contrast thresholds were roughly unaffected by the noise *spatial* window as similar contrast thresholds were observed in spatially localized and extended noise both when the temporal window was localized (SL-TL and SE-TL) and extended (SL-TE and SE-TE). This was statistically validated by a  $2 \times 2 \times 2$  ANOVA (task  $\times$  spatial window  $\times$  temporal window), which revealed no significant effect of spatial window [ $F(1,3) = 1.83, p = 0.27$ ] and



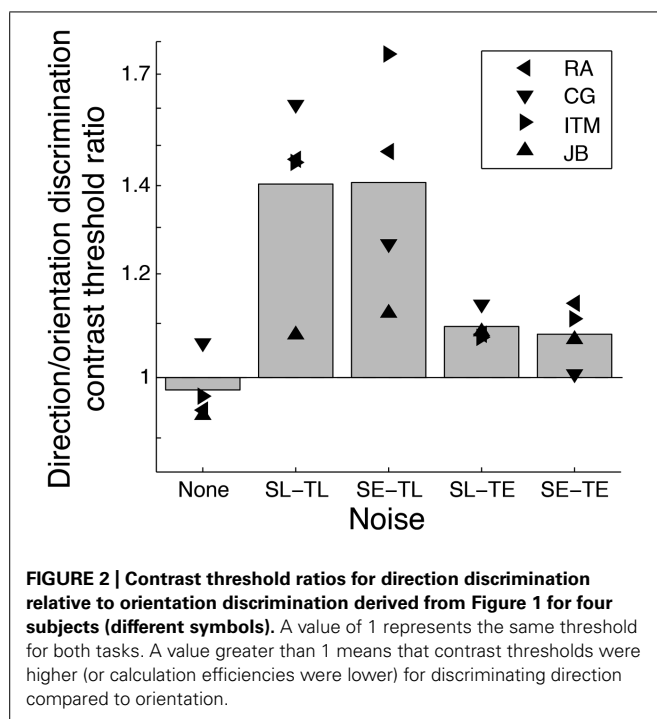
**FIGURE 1 | Contrast thresholds obtained for four subjects (different symbols) in five different noise conditions for orientation (open symbols) and direction (closed symbols) discrimination.** In the four noise conditions, the noise was either spatially localized (SL) or extended (SE) and temporally localized (TL) or extended (TE). Each data point corresponds to the average of four staircases. For clarity, the standard error are not shown, but were all smaller than 0.06 log units (i.e., less than a factor of 1.15).

no task  $\times$  spatial window interaction [ $F(1,3) = 0.019, p = 0.90$ ]. On the other hand, contrast thresholds varied with the noise *temporal* window [ $F(1,3) = 57.9, p < 0.01$ ] and varied differently for the two tasks (task  $\times$  temporal window interaction,  $F(1,3) = 10.4, p < 0.05$ ). Specifically, contrast thresholds were lower (i.e., higher calculation efficiency) in temporally localized noise (SL-TL and SE-TL) relative to temporally extended noise (SL-TE and SE-TE, respectively) by a factor of about 2 for orientation discrimination and 1.4 for direction discrimination.

**Figure 2** illustrates the contrast threshold ratios for direction relative to orientation discrimination represented in **Figure 1**. Similar contrast thresholds were observed for orientation and direction discrimination (i.e., ratios close to 1) in absence of noise and in temporally extended noise (SL-TE and SE-TE), but contrast thresholds were substantially better (by a factor of about 1.4 on average) for orientation than for direction discrimination in temporally localized noise (SL-TL and SE-TL, respectively).

## DISCUSSION

Calculation efficiency ratios (which can be directly inferred from contrast threshold ratios in high noise) of direction discrimination relative to orientation discrimination varied with the noise temporal window: a substantial difference was observed in temporally localized noise (threshold ratio of  $\sim 1.4$ ), but not in temporally extended noise (ratio close to 1). The purpose of external noise paradigms is generally to estimate the calculation efficiency in absence of noise by assuming that it is the same as the measured calculation efficiency in high noise. However, the fact that the calculation efficiency ratios varied with the noise temporal window implies that in at least one condition the measured calculation



efficiency in high noise did not correspond to the calculation efficiency in absence of noise. Indeed, the calculation efficiency in absence of noise cannot both differ substantially for orientation and direction discrimination as measured in localized noise and be similar for orientation and direction discrimination as measured in extended noise. Thus, in at least one condition, the calculation efficiency measured in high noise did not correspond to the calculation efficiency in absence of noise, which violates the noise-invariant processing assumption and compromises the application of the external noise paradigm.

In absence of noise (i.e., in internal noise), no substantial contrast threshold difference was observed (ratios close to 1) as in temporally extended noise. Given that internal noise is expected to be temporally extended (it does not turn on and off with the stimulus) and that contrast thresholds were similar for orientation and direction discrimination as in extended noise, this suggests that the calculation efficiencies in absence of noise did not differ between tasks. As a result, there was no evidence of a violation of the noise-invariant processing assumption when using temporally extended noise so the calculation efficiency measured in temporally extended noise likely reflects the calculation efficiency in absence of noise. Contrariwise, the facts that internal noise is not temporally localized and that a different pattern of results was observed in temporally localized noise suggest that the calculation efficiencies measured in temporally localized noise were not both the same as the calculation efficiencies in absence of noise. This shows a violation of the noise-invariant processing assumption, as the measured calculation efficiency in localized noise cannot be assumed to be the same as the calculation efficiency in absence of noise.

The results of the current study suggest that when temporally localized noise dominated the most sensitive detectors were labeled

for orientation only (e.g., simple cells), whereas when temporally extended noise dominated (which includes internal noise) the most sensitive detectors were labeled for both direction and orientation (e.g., direction-selective complex cells). Thus, which detectors were the most sensitive depended on the temporal window of the dominating noise source. This suggests that temporally localized noise impaired more the sensitivity of detectors labeled for orientation and direction (e.g., direction-selective complex cells, which would be the most sensitive in absence of noise) than the ones labeled for orientation only (e.g., simple cells). This greater masking for direction selective detectors can be explained by the sharp contrast transient onset and/or offset of the noise. Note that technically, there is more luminance transient between two different noise frames than between a mean gray frame and a noise frame. However, the temporal envelope of the localized noise contains a strong transient (turns on and off, i.e., noise contrast varies from 0 to high to 0) whereas the extended noise does not (it is continuously present, i.e., constant mean contrast). This corresponds to the subjective impression: temporally localized noise suddenly appears causing a sharp transition from a blank to a noisy display whereas temporally extended noise appears to be constantly displayed even if it is dynamic. Thus, the current results suggest that the sharp transients of the noise envelope (i.e., noise onset and offset) impair more the detectors labeled for both orientation and direction than the ones labeled for orientation only.

Given that transients caused by localized noise cause additional masking, one could expect thresholds to be lower (i.e., better) in extended noise than in localized noise, which is opposite to the current findings (Figure 1). Even though adaptation is known to reduce responsiveness of stimulated cells (Giaschi et al., 1993), it is unlikely that it affects contrast threshold in high noise because adaptation would affect the responses related to both the signal and noise leaving the signal-to-noise ratio intact. This would have no impact on contrast threshold given that contrast threshold in high noise is proportional to the noise contrast (Pelli, 1981). Conversely, there are at least two reasons why extended noise could have a greater masking effect than localized noise. First, the visual system has a limited temporal resolution and therefore integrates some noise outside the signal temporal window (i.e., just before the target onset and after the target offset). Second, localized noise has the advantage of reducing temporal uncertainty, which is obviously not the case for temporally extended noise. Thus, adding noise outside the temporal window of the signal (i.e., passing from localized to extended noise) can facilitate contrast threshold by removing noise onset and offset transient, but impair contrast threshold by introducing more noise and increasing temporal uncertainty. It is therefore not surprising that contrast thresholds in temporally extended noise are higher than in temporally localized noise even though there is no noise onset and offset transient in extended noise.

By compromising the estimation of the calculation efficiency in absence of noise, a violation of the noise-invariant processing assumption also compromises the estimate of the internal equivalent noise. Based on the linear amplifier model (Pelli, 1981; Pelli and Farell, 1999), contrast threshold in absence of noise depends on both internal equivalent noise and calculation efficiency. By

knowing the contrast threshold in absence of noise and by assuming that the calculation efficiency in absence of noise is the same as the measured calculation efficiency in high noise, the internal equivalent noise can be calculated. If the calculation efficiency in absence of noise cannot be assumed to be the same as the measured calculation efficiency in high noise, then the internal equivalent noise cannot be calculated. For instance, Arena *et al.* (2013) observed that aging affected contrast thresholds in low, but not in high, localized noise. Given that contrast thresholds in high noise depend only on the calculation efficiency and not on the internal equivalent noise, they concluded that the calculation efficiency in low noise was not affected with aging and therefore attributed the age-related sensitivity losses in low noise to an increase in internal equivalent noise. However, given that the measured calculation efficiency in absence of noise does not correspond to the measured calculation efficiency in high localized noise (as suggested by the current findings), both the calculation efficiency in absence of noise and the internal equivalent noise remains unknown and it is not possible to determine whether the age-related sensitivity loss in low noise was due to a lower calculation efficiency or higher internal equivalent noise.

The current study found that the most sensitive detectors underlying motion processing varied with the noise temporal window. In temporally extended noise (which includes internal noise), the most sensitive detectors were labeled for both orientation and direction, whereas in temporally localized noise, they were labeled for orientation, but not direction. In absence of noise (i.e., in internal noise), the most sensitive detectors would be labeled for both orientation and direction, which suggests, as expected, that internal noise limiting motion processing is temporally extended. Thus, to characterize motion processing in absence of noise, such as measuring internal equivalent noise and calculation efficiency, external noise should be temporally extended to avoid violating the noise-invariant processing assumption.

## ACKNOWLEDGMENTS

This research was supported by NSERC discovery fund to Jocelyn Faubert and Essilor International.

## REFERENCES

- Allard, R., and Cavanagh, P. (2011). Crowding in a detection task: external noise triggers change in processing strategy. *Vision Res.* 51, 408–416. doi: 10.1016/j.visres.2010.12.008
- Allard, R., and Faubert, J. (2008). The noisy-bit method for digital displays: converting a 256 luminance resolution into a continuous resolution. *Behav. Res. Methods* 40, 735–743. doi: 10.3758/BRM.40.3.735
- Allard, R., and Faubert, J. (2013). Zero-dimensional noise is not suitable for characterizing processing properties of detection mechanisms. *J. Vis.* 13, 1–3. doi: 10.1167/13.10.25
- Allard, R., Renaud, J., Molinatti, S., and Faubert, J. (2013). Contrast sensitivity, healthy aging and noise. *Vision Res.* 92, 47–52. doi: 10.1016/j.visres.2013.09.004
- Arena, A., Hutchinson, C. V., Shimozaki, S. S., and Long, M. D. (2013). Visual discrimination in noise: behavioural correlates of age-related cortical decline. *Behav. Brain Res.* 243, 102–108. doi: 10.1016/j.bbr.2012.12.039 doi: 10.1016/j.bbr.2012.12.039
- Giaschi, D., Douglas, R., Marlin, S., and Cynader, M. (1993). The time course of direction-selective adaptation in simple and complex cells in cat striate cortex. *J. Neurophysiol.* 70, 2024–2034.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49(Suppl. 2), 467+. doi: 10.1121/1.1912375
- Pelli, D. G. (1981). *The Effects of Visual Noise*. Ph.D. thesis, Cambridge University, Cambridge.
- Pelli, D. G., and Farell, B. (1999). Why use noise? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 16, 647–653. doi: 10.1364/JOSAA.16.000647
- Watson, A. B., and Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Res.* 21, 1115–1122. doi: 10.1016/0042-6989(81)90014-6

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 January 2014; paper pending published: 03 March 2014; accepted: 22 April 2014; published online: 15 May 2014.

Citation: Allard R and Faubert J (2014) Motion processing: the most sensitive detectors differ in temporally localized and extended noise. *Front. Psychol.* 5:426. doi: 10.3389/fpsyg.2014.00426

This article was submitted to *Perception Science*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Allard and Faubert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Distinct mechanisms subserve location- and object-based visual attention

Wei-Lun Chou<sup>1,2</sup>, Su-Ling Yeh<sup>1,3</sup> \* and Chien-Chung Chen<sup>1,3</sup>

<sup>1</sup> Department of Psychology, National Taiwan University, Taipei, Taiwan

<sup>2</sup> Department of Psychology, Fo Guang University, Yilan, Taiwan

<sup>3</sup> Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan

## Edited by:

Jocelyn Faubert, Université de Montréal, Canada

## Reviewed by:

Zhong-Lin Lu, University of Southern California, USA

Jason M. Gold, Indiana University Bloomington, USA

## \*Correspondence:

Su-Ling Yeh, Department of Psychology, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan  
e-mail: [suling@ntu.edu.tw](mailto:suling@ntu.edu.tw)

Visual attention can be allocated to either a location or an object, named location- or object-based attention, respectively. Despite the burgeoning evidence in support of the existence of two kinds of attention, little is known about their underlying mechanisms in terms of whether they are achieved by enhancing signal strength or excluding external noises. We adopted the noise-masking paradigm in conjunction with the double-rectangle method to probe the mechanisms of location-based attention and object-based attention. Two rectangles were shown, and one end of one rectangle was cued, followed by the target appearing at (a) the cued location; (b) the uncued end of the cued rectangle; and (c) the equal-distant end of the uncued rectangle. Observers were required to detect the target that was superimposed at different levels of noise contrast. We explored how attention affects performance by assessing the *threshold versus external noise contrast (TvC) functions* and fitted them with a divisive inhibition model. Results show that location-based attention – lower threshold at cued location than at uncued location – was observed at all noise levels, a signature of signal enhancement. However, object-based attention – lower threshold at the uncued end of the *cued* than at the *uncued* rectangle – was found only in high-noise conditions, a signature of noise exclusion. Findings here shed a new insight into the current theories of object-based attention.

**Keywords:** attention mechanisms, location-based attention, object-based attention, threshold versus external noise contrast (TvC) function, noise-masking paradigm, divisive inhibition model

Our visual world is full of information; however, not all can be selected for further processing due to limited capacity. Mechanisms of attention are thus employed to prioritize the processing of particular information. Past studies have shown that visual attention can be allocated either to a spatial location or to an object, called location-based attention or object-based attention, respectively (Posner, 1980; Duncan, 1984; Tipper et al., 1991; Egly et al., 1994; Gibson and Egeth, 1994; Brawn and Snowden, 2000).

In a seminal work, Egly et al. (1994) used a double-rectangle display to demonstrate both location-based attention and object-based attention. They presented two outlined rectangles, with one end of one rectangle brightened as a cue to indicate the possible location of a target. The target was a small solid square, shown subsequently within one end of a rectangle. Location-based attention was indicated by the *spatial-cueing* effect: reaction times (RTs) were shorter when the target appeared at the cued location than the uncued location. Object-based attention was indicated by the *same-object advantage*: RTs were shorter when the target appeared at the uncued end of the *cued* rectangle than at the *uncued* rectangle, with an equal cue-to-target distance between the two. Concurring with Egly et al. (1994), a series of studies using various stimuli and tasks have demonstrated the spatial-cueing effect and the same-object advantage (Moore et al., 1998; Abrams and Law, 2000; Lamy and Tsal, 2000; Moore and Fulton, 2005; Brown et al.,

2006; Matsukura and Vecera, 2006; Shomstein and Behrmann, 2008).

The spatial-cueing effect has been explained by the movement of attention from one location to another in visual space. On valid trials, a shift of attention can be initiated to the expected target *location* before the target appears, thereby producing an RT or accuracy benefit (Posner, 1980). On the two kinds of invalid trials, however, a shift of attention would be initiated to a location on the wrong site of the display from the actual target location. This would produce an RT or accuracy cost because attention would need to be realigned with the correct target location after the target's appearance.

The same-object advantage has been explained mainly by two competing theories. The *spreading hypothesis* states that when attention is cued to a location within an object, attention will spread automatically from the cued location to the whole object (e.g., Davis and Driver, 1997; Kasai and Kondo, 1997; Richard et al., 2008). Such spread of attention explains the participants' better visual performance when the target was shown on the cued object than on the uncued object. Since the attentional modulation is triggered by a location cue and spreads to the whole object, the same-object advantage should be an instance of location-based attention. That is, the underlying mechanism of object-based attention is the same as that of location-based attention. In addition, it is shown that improvement of visual performance in a location-based attention task can be due to (a) the participant



being more sensitive to a target at the cued location than that at the uncued one; and/or (b) the participant being less influenced by irrelevant visual information (Lu and Doshier, 1998). Hence, these two factors should be able to account for object-based attention as well, if it shares the same mechanism as location-based attention.

On the other hand, the *prioritization hypothesis* (Shomstein and Yantis, 2002) suggests that object-based attention reflects a specific attentional prioritization strategy rather than the modulation of an early sensory enhancement extending from the location-based attention. That is, the prioritization hypothesis does not take any position regarding the similarity of the mechanisms between location- and object-based attention. At best, it would predict different mechanisms for the *exogenous* spatial-cueing effect and the strategically object-based scanning strategy. Therefore, the same-object advantage cannot be explained by a change in early sensory mechanisms.

Here, we are interested in the mechanisms that subserve location- and object-based attention, especially whether the mechanisms underlying these two types of attention are the same. Notice that previous investigations adopting the double-rectangle method generally used RT measurement with a single level of task difficulty (Egley et al., 1994; Moore et al., 1998; Abrams and Law, 2000; Lamy and Tsal, 2000; Moore and Fulton, 2005; Brown et al., 2006; Shomstein and Behrmann, 2008). RT measurement may reflect processing speed, response bias, or a combination of the two (Ratcliff, 1978), making it hard to infer the underlying mechanisms. In addition, while an estimation of response variability is important to evaluate certain theories of location-based attention (Lu and Doshier, 1998), it is difficult to separate measurement error from the experimental procedure and the variability of the internal responses in the RT measurement.

We used a noise-masking paradigm (Nagaraja, 1964; Legge et al., 1987; Pelli, 1991; Lu and Doshier, 1998) that can evaluate the variability in the response of the visual system in the double-rectangle display to probe the mechanism(s) of location-based attention and object-based attention. In a typical noise-masking paradigm, the task of the observer is to detect a pre-designated target that is superimposed on a patch of white noise. In the context of our experiment, the target was a periodic pattern defined by a Gabor function, which is a product of a sine wave and a Gaussian envelope, while the noise was a random modulation of luminance. The intensity of the noise mask was defined by contrast, or the theoretical half range of the luminance modulation defined by a uniform distribution divided by the mean luminance. By systematically measuring the target threshold at different external noise levels, we can measure the *threshold versus external noise contrast (TvC) functions*. With an appropriate model, this information allows an estimation of the response properties and variability of the target detection mechanisms, thus providing a more comprehensive estimation of various perceptual mechanisms (Nagaraja, 1964; Legge et al., 1987; Pelli, 1991; Lu and Doshier, 1998; Chen and Tyler, 2001; Wu and Chen, 2010).

By taking advantage of the double-rectangle method, we evaluated the TvC functions of attended and unattended location/object

within a single paradigm. In a two-alternative intervals choice task (Figure 1), participants were asked to detect a Gabor target that was superimposed on a noise pattern. The displays, if not stated otherwise, consisted of two vertical rectangles that were presented on each side of fixation. The four ends of the rectangles were where the cue (or target) was likely to occur. The target could occur at one of the three possible locations: the cued location (*valid*), the uncued location but on the cued object (*same-object*), or an equidistant location on the uncued object (*different-object*). Then, we measured the TvC functions for all the different conditions so that we can compare location-based attention and object-based attention and infer their mechanisms directly. If their mechanisms are identical, they should show the same kind of shift in the TvC functions.

## MATERIALS AND METHODS

### ETHICS STATEMENT

The use of human participants was approved by the IRB of National Taiwan University Hospital and followed the guideline of Helsinki Declaration. The written informed consent was obtained from each participant.

### APPARATUS

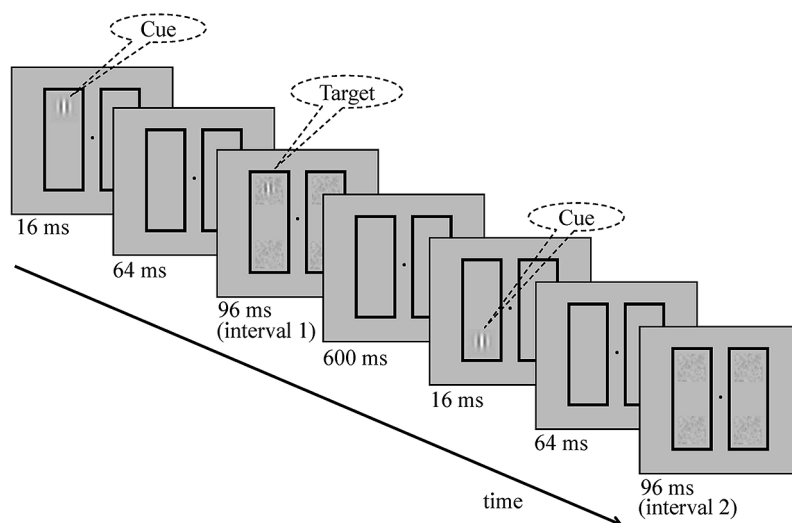
Two ViewSonic (15") CRT monitors, each driven by a Radeon 7200 graphic board, were used to present the stimuli. The graphic board provided 10-bit digital-to-analog converter depth and was controlled by a Macintosh computer. A beam splitter was used to combine lights from the two CRT monitors. The target was presented on one monitor and the cue and the external noise patch (mask) on the other. This two-monitor setup had the advantage that the contrast of the target could be controlled independently while keeping the context (the cue and the mask) identical in two intervals of a trial. At a viewing distance of 128 cm, the resolution on a 640 × 480 pixels monitor was 60 pixels per degree. The refresh rate of the monitors was 66 Hz. The viewing field was 10.7° × 8° (horizontal × vertical), and the mean luminance of the displays was 74.9 cd/m<sup>2</sup>. The LightMouse photometer (Tyler and McBride, 1997) was used to measure the full-detailed input-output intensity function of the monitors, and this information was then used to compute linear lookup table settings so as to linearize the output within 0.2%.

### STIMULI AND DISPLAY

Figure 1 illustrates the stimuli and sequence of events for a trial. The displays are comprised of a pair of adjacent vertical rectangles. The fixation was a small dot. Each rectangle (1.63° × 4.88°, with a stroke width 0.13°) was centered 3° from fixation. The cue and the target were vertical Gabor patches defined by the following equation:

$$G(x, y, c, u_x, u_y) = L + L * c * \cos(2\pi f x) * \exp\left(-\frac{(x - u_x)^2}{2\sigma^2}\right) * \exp\left(-\frac{(y - u_y)^2}{2\sigma^2}\right),$$

where  $L$  was the mean luminance,  $c$  was the contrast ranging from 0 to 1,  $f$  was the spatial frequency,  $\sigma$  was the scale parameter of the Gaussian envelope,  $u_x$  was the horizontal displacement, and



**FIGURE 1 | Schematic overview of a typical valid trial with the target showing in interval 1.** The task was to detect the target (a Gabor patch) superimposed on different levels of noise (mask) contrast in a two-alternative forced-choice paradigm. In each interval, a cue was flashed

first for 16 ms, followed by a 64 ms blank, and then a stimulus presentation (either target-plus-noise mask or noise mask alone). Two intervals were separated by a 600 ms blank. The rectangles and the fixation point were always on-screen.

$u_y$  was the vertical displacement. Both Gabor patches had a spatial frequency ( $f$ ) of 1.3 cycles/deg and a scale parameter ( $\sigma$ ) of 0.3536°. The contrast of the cue ( $c$ ) was  $-6$  dB or 50%. For each external noise frame the pixel gray-levels were sampled from a Gaussian distribution.

## PROCEDURE

A two-alternative forced-choice paradigm was used to measure the threshold of the target (Figure 1). The cue was presented at one of four possible locations in each interval. After that, the target was presented at one of the three possible locations: (1) the cued location (*valid* trials), (2) the uncued end within the cued object (*same-object* trials), or (3) the uncued end within the uncued object (*different-object* trials) in one of the intervals.

A fixation display (a central fixation point and two outline rectangles) was presented first, followed by a 16-ms cue display, then a 64-ms fixation display, and finally a 96-ms target display (a target and four mask patches). The stimulus onset asynchrony between the cue and the target was 80 ms, the inter-stimulus-interval within a trial was 600 ms, and the inter-trial-interval was 800 ms. At the beginning of each trial an audio tone was presented as a signal to start. Correct and incorrect responses were followed by auditory feedbacks.

Each block of seven external noise levels ( $-\infty$ ,  $-26$ ,  $-22$ ,  $-18$ ,  $-14$ ,  $-10$ ,  $-6$  dB) were presented in random order, and each block contained the three attention conditions (valid, same-object, and different-object). The threshold was defined at 75% correct response level, measured by the PSI threshold-seeking algorithm (Kontsevich and Tyler, 1999). For each threshold measurement, two practice trials preceded 40 formal trials. Within a single block, four thresholds were measured in an interleaved way – two for the valid condition, one for the same-object and one for the different-object conditions, making the total number of valid trials (84

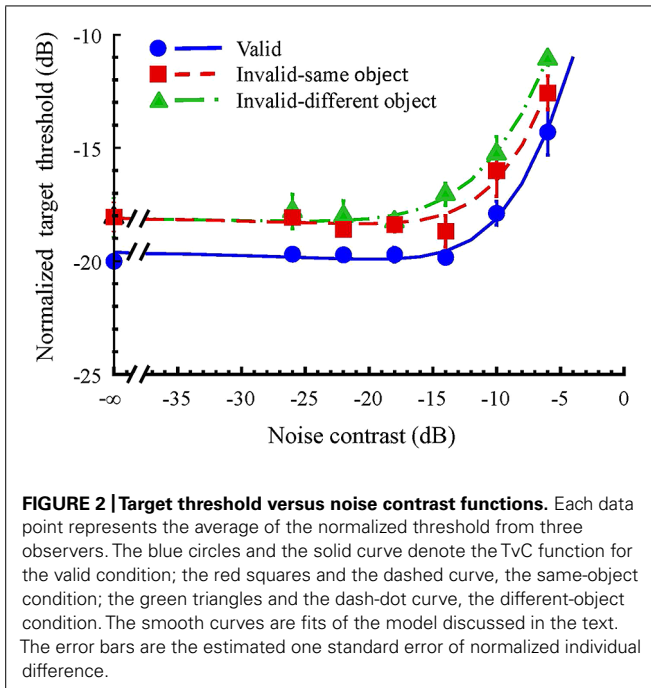
trials) twice as many as that of the same-object or different-object trials (42 trials). That is, the cue validity for predicting the target location was 50%. The sequence of trials was pseudo-randomized. The TvC function of the valid condition is the average of two threshold measurements. Each data point reported was an average of four to eight repeated measures. The task was to indicate which interval contained the target by pressing a corresponding key. Participants were told that the two outline rectangles were task-irrelevant, and they were well informed about the cue-target relationship.

## PARTICIPANTS

Three participants with normal or corrected-to-normal visual acuity were tested. RY and TH were naïve as to the purposes of this study and WL was one of the authors.

## RESULTS

Figure 2 shows the result averaged across three participants. The blue circles and solid curve denote the TvC function for the valid condition; red squares and dash curve, the same-object condition; and green triangles and dash-dot curve, the different-object condition. To account for the individual difference in overall sensitivity to the target, we scaled each threshold by that measured at zero noise contrast of the valid condition of the corresponding participant before averaging. When there was no noise mask, the threshold for the valid condition was lower than that for both invalid conditions. The difference was 2 dB [ $t(2) = 3.46$ ,  $p = 0.037 < 0.05$ ] between the valid cue and both the invalid conditions. Such difference between the valid and invalid conditions remained as the mask increased. Thus, the TvC functions of the invalid conditions look like a vertically shifted copy of the valid condition on log-log coordinates. Such general facilitation on target detection suggests



that the effect of the valid cue was to increase the sensitivity to the target (Cohn and Lasley, 1974; Lu and Doshier, 1998; Zenger et al., 2000; Pestilli and Carrasco, 2005; Chen and Tyler, 2010).

The target detection thresholds were not influenced by the low contrast noise mask for all attention conditions. As a result, all TvC functions were flat at low noise contrasts. When the noise contrast reached a critical value, the threshold began to increase with noise contrast. Here, whether or not the cue and the target were within the boundary of an object had an effect. The threshold increment for the different-object condition started at a lower noise contrast than that for the same-object condition. As a result, the TvC function for the different-object condition showed a leftward shift from the TvC function for the same-object condition. This suggests that the noise effect on target detection in the same-object condition is different from that in the different-object condition.

Our result cannot be explained by an inter-hemispherical effect. In a control condition, we used horizontal rectangles as the objects. We measured the target threshold at noise level  $-\infty$  and  $-6$  dB. There was no statistical significant difference [ $t(11) = -1.1$ ,  $p = 0.30$ ] in target threshold between the vertical and the horizontal object configurations, averaged across all conditions and observers.

## MODEL

We fitted the TvC functions by a version of the divisive inhibition model (Ross and Speed, 1991; Wilson and Humanski, 1993; Foley, 1994; Teo and Heeger, 1994; Watson and Solomon, 1997; Snowden and Hammett, 1998; Chen and Foley, 2004) modified to account for the noise-masking experiment (Lu and Doshier, 1998; Goris et al., 2008; Chen and Tyler, 2010). This model integrates features from the divisive inhibition models for pattern

detection and discrimination (Foley, 1994; Chen and Foley, 2004) and conventional models for noise masking (e.g., Lu and Doshier, 1998). Chen and Tyler (2010) used a similar model to account for the cueing effect in a noise-masking paradigm. **Figure 3** shows a diagram of this model. There are several stages in this model. The first stage is a band of linear filters operating on the input images. The excitation of a linear filter is then half-wave rectified, raised to a power and scaled by a divisive inhibition input to form the response of the target detector. The decision variable is the ratio of the response of the target detector and the noise from different sources.

Each mechanism  $j$  contains a linear operator within a spatial sensitivity profile  $f_j(x, y)$ . The excitation of this linear operator to the  $i$ -th image component  $g_i(x, y)$  is specified as:

$$E_{ji}' = \sum_x \sum_y f_j(x, y) g_i(x, y) \quad (1)$$

where the linear filter  $f_j(x, y)$  is defined by a Gabor function (see "Materials and Methods"). Suppose that the image component  $g_i(x, y)$  has a contrast  $C_i$ . Summing over  $x$  and  $y$ , Eq. (1) can be simplified to

$$E_{ji}' = Se_{ji} C_i \quad (1')$$

where  $Se_{ji}$  is a constant defining the excitatory sensitivity of the mechanism to the stimulus ( $j = t$  for the target and  $j = m$  for the mask). Detailed derivation of Eq. (1)' from Eq. (1) has been discussed elsewhere (Chen and Tyler, 1999; Chen et al., 2000).

The excitation of the linear operator is half-wave rectified (Foley, 1994; Teo and Heeger, 1994; Foley and Chen, 1999) to produce the rectified excitation  $E_{ji}$

$$E_{ji} = \max(E_{ji}', 0) \quad (2)$$

where  $\max$  denotes the operation of choosing the greater of the two numbers.

The total excitation of the  $j$ -th mechanism  $E_j$  is the sum of excitations produced by all image components. The response of the  $j$ -th detector is then  $E_j$ , raised by a power  $p$  and divided by a divisive inhibition term  $I_j$  plus an additive constant  $z$ . That is,

$$R_j = E_j^p / (I_j + z) \quad (3)$$

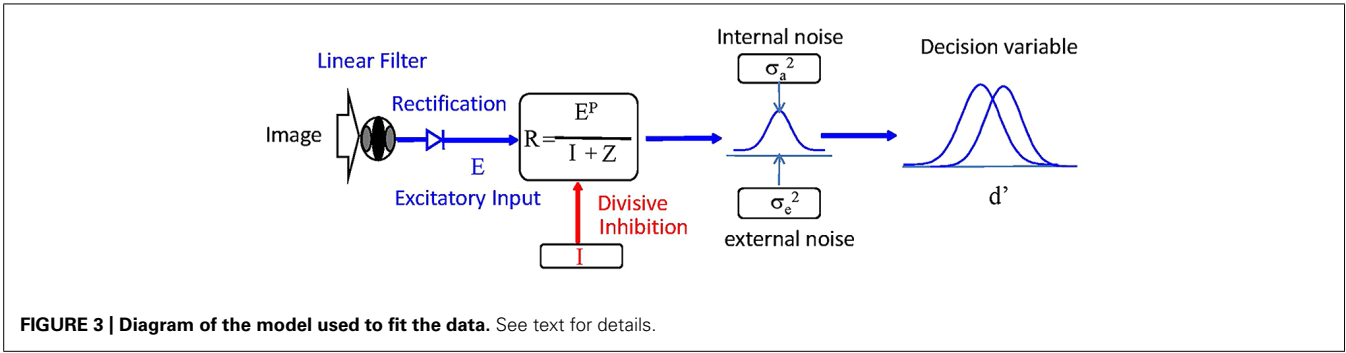
where  $I_j$  is the summation of a non-linear combination of the excitations of all relevant mechanisms. This divisive inhibition term  $I_j$  can be represented as

$$I_j = \sum_i (S_{ji} C_i)^q \quad (4)$$

where  $S_{ji}$  is the weight of the contribution from each component to the inhibition term.

The contribution of a detector to the visual performance is limited by the noise. We consider two sources of noise in this model: the internal noise inherent in the system, and the external noise provided by the noise patterns. The variability produced by the internal noise,  $\sigma_a^2$ , is a constant for all detectors in the model. The variability produced by the external noise,  $\sigma_e^2$ , is proportional to the square of the contrast noise mask; that is,

$$\sigma_e^2 = w_m C_m^2 \quad (5)$$



where  $w_m$  is a scalar constant that determines the amount of contribution of the noise mask to the variance of the response. Pooling the effects of these two noise sources, the variance of the response distribution in each detector is

$$\sigma_r^2 = (\sigma_a^2 + \sigma_e^2) \tag{6}$$

In the context of our experiment, the observer compared the response to the stimuli in both intervals at the three possible target locations. The observer can detect the target if the difference between the response to the target + mask,  $R_{j, t+m}$ , and that to the mask alone,  $R_{j,m}$ , is greater in at least one channel than is the limitation imposed by the noise. In practice, we need to consider only the mechanism that produces the greatest response difference between the target + mask and the mask alone conditions. Thus, we can drop the subscript  $j$  for this study. That is, the decision variable  $d'$  is,

$$d' = (R_{m+t} - R_m) / (2\sigma_r^2)^{1/2} \tag{7}$$

The threshold is defined when  $d'$  reaches unity.

**Table 1** shows the parameter of the model. To reduce the mathematical redundancy in the model, we fixed the sensitivity to the target,  $Se_t$ , for the valid cue condition to be 100 and the size of the internal noise,  $\sigma_a^2$  to be 1. As shown in the Results section, the TvC functions for the invalid conditions are vertically shifted

copies of the valid condition on log–log coordinates. As shown in **Figure 4A**, such vertical shift of TvC functions can be achieved by changing the sensitivity to the target,  $Se_t$ . Hence, our data suggest that the sensitivity to the target to be different for the valid and invalid cue conditions. This result is consistent with the models proposed by Reynolds and Heeger (2009), which suggested that spatial attention can operate in the early visual areas by affecting the attention field, and by Lu and Doshier (1998), which suggested that spatial cue enhances the target signal.

The TvC function for the different-object condition shifted to the left from that of the same-object condition. Such horizontal shift can be implemented a change in the relative contribution of the external noise  $w_m$  (**Figure 4B**). Thus, our result suggests that the contribution of the external noise to the response variance,  $w_m$ , is different in the same-object and the different-object conditions. Notice that in the valid condition, the target and the cue were also presented within the boundary of the same object. Therefore, we constrained all parameters to be the same across conditions except for sensitivity to the target,  $Se_t$ , and the contribution of the external noise,  $w_m$ . This model fits the data well; the root of mean squared error (RMSE) was 0.27. This model explains 98.61% of all variance in the averaged data.

To further validate our interpretation of the data, we tried various constraints to the model. If we constrained the sensitivity to the target,  $Se_t$ , to be the same for all conditions, the sum of squared error (SSE) of the model increased significantly [ $F(1,12) = 73.82$ ,  $p < 0.0001$ ] even when we took the number of free parameters into account. Similarly, constraining the contribution of the external noise,  $w_m$ , to be the same for both invalid conditions significantly increased the SSE [ $F(1,12) = 16.63$ ,  $p < 0.05$ ]. Therefore, the change of sensitivity to the target is necessary to explain the spatial-cueing effect while the change of the contribution of the external noise is necessary to explain the same-object advantage.

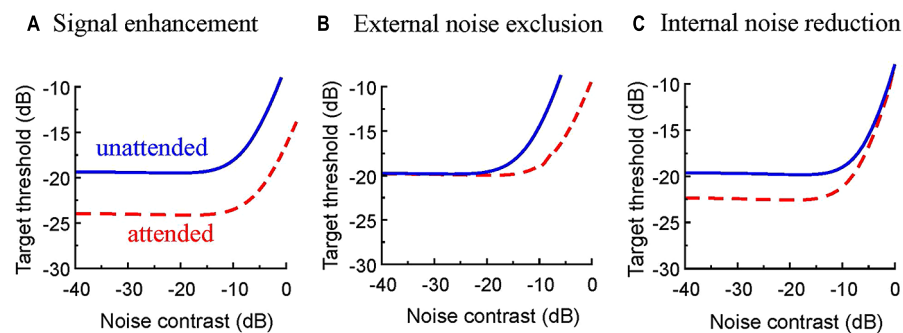
Lu and Doshier (1998) suggested a mechanism of internal noise reduction for attention. That is, the effect of the cue is to reduce the effect of the additive noise in the system. In our model, this can be implemented by changing the value of the internal noise parameter  $\sigma_a$ . As shown in **Figure 4C**, such change in parameter value will cause TvC function to shift vertically in the low noise contrasts. However, the TvC function would merge together at high contrasts. We did not find such a trend in our data. Hence, our result cannot be explained by a reduction of additive internal noise. We also found that more free parameters in the model never

**Table 1 | The estimated parameters of the model.**

	Conditions		
	Valid	Same-object	Different-object
$Se_m$	2.47	2.47	2.47
$Se_t$	100*	93.99	93.99
$Si_t$	308.75	308.75	308.75
$z$	1.62	1.62	1.62
$w_m$	5.71	5.71	11.48
$\sigma_a^2$	1*	1*	1*
$p$	3.11	3.11	3.11
$q$	2*	2*	2*

\*Fixed value, not a free parameter.





**FIGURE 4 | Performance signatures in threshold contrast versus external noise contrast (TvC) functions.** (A) If the TvC functions are a vertically shifted copy of each other, that is, the same target would have different thresholds in the attended and unattended conditions, this suggests that the participant has a different sensitivity to the target in the two conditions. Hence, the effect of attention is to enhance the sensitivity [Se in Eq. (1)] to the target in our model. (B) Suppose that the TvC functions for the attended condition is a rightward-shifted copy of the unattended condition. It means

that the same external noise level can have different effects on target detection in the attended and unattended conditions. This suggests that attention allows the participants to exclude noise in the stimuli more easily. This corresponds to a reduction of the contribution from the external noise [ $w_m$  in Eq. (5)] in our model. (C) If the TvC functions showed a vertical shift at low noise contrast, but merged at high contrasts, the effect of attention is to reduce the internal noise. This corresponds to a reduction of the internal noise parameter [ $\sigma_2^2$  in Eq. (6)] in our model.

produced a significant improvement of goodness-of-fit. Thus, no extra factors are necessary to explain our results.

## DISCUSSION

The current study systematically probed the target threshold improvement by location- and object-based attention with different noise levels using the double-rectangle method, and the results suggest that location- and object-based attention involve different mechanisms. Location-based attention operates by enhancing signal strength, whereas object-based attention operates by excluding external noise. This study is the first to demonstrate the discrepancy in the TvC functions of location- and object-based attention within a single task.

In previous studies, location- and object-based attention were examined separately by the noise-masking paradigm. Location-based attention was observed in both no-noise and high-noise conditions (Doshier and Lu, 2000; Lu and Doshier, 2000), consistent with our results. However, Han et al. (2003) found that object-based attention was also observed in both no-noise and high-noise conditions, inconsistent with our findings here. Notice that Han et al. (2003) compared the performances of tasks that required participants to attend to only *one* object versus *two* spatially separated objects. Object-based attention was indexed by higher accuracy of reporting two attributes belonging to a single object than different objects, and it was shown in both no- and high-contrast noise conditions in Han et al.'s (2003) study. It is reasonable to argue that their participants may have changed their attentional window – like a zoom lens (Eriksen and Yeh, 1985) – from “wide” in the two-object condition to “small” in the single-object condition. Accordingly, the differences between the two-object and single-object conditions not only are the number of attended objects but also the size of spatial attention (Davis et al., 2000).

This argument is supported by Liu et al. (2009) with a design identical to Han et al.'s (2003). The magnitude of the same-object advantage was modulated by the required precision of judgments:

the higher the task precision, the larger the difference in performance between the two-object and the single-object conditions (Liu et al., 2009). Assuming that attentional window is wide in the two-object condition, the density of attentional resource should be low due to the reciprocal relationship between size and density of attentional distribution (Eriksen and St. James, 1986; LaBerge and Brown, 1989). The low-precision task that requires less resources can be performed equally well with less attentional resource in the two-object condition as opposed to the one-object condition – leading to reduced or no same-object advantage. The critical comparison in their study – two-object and single-object conditions – may not reflect object-based attention but rather a change in the window size of spatial attention. Indeed, the modulation pattern of “object-based” attention in Han et al.'s (2003) study is similar to the modulation pattern of location-based attention (Doshier and Lu, 2000; Lu and Doshier, 2000): both can be observed in no-noise and high-noise conditions. However, the double-rectangle method compares the same-object and different-object conditions based on an equal cue-to-target distance between the two conditions. Using the double-rectangle method, we rule out the confounding of location-based attention in the current study and find that object-based attention is observed only in high-noise conditions, indicating that external noise exclusion plays a critical role in object-based attention.

The qualitative difference between the intrinsic mechanisms of location-based and object-based attention suggests that object-based attention is not an outcome of the spreading from the location-based attention, which is a finding arguing against the well-accepted *spreading hypothesis* (e.g., Davis and Driver, 1997; Kasai and Kondo, 1997; Richard et al., 2008). Instead, we suggest that object-based attention reflects a qualitatively different kind of attentional orienting that is independent of location-based attention, rather than the modulation of an early sensory enhancement extending from location-based attention. This argument is also against the *prioritization hypothesis* proposed by Shomstein and Yantis (2002), who claimed that object-based attention



reflected strategic prioritization regardless of location-based effect and that neither was it due to object-based perceptual enhancement. However, using the noise-masking paradigm, we provide evidence for the underlying mechanism of object-based attention. The current finding of the leftward-shifted copies of the TvC functions in the same-object and different-object conditions suggests that the underlying mechanism of object-based attention is to exclude external noise, an evidence of object-based perceptual enhancement.

In our experiment, the target may appear in one of the three possible locations. As a result, the participant would experience a greater uncertainty in the invalid conditions, in which the participant needed to monitor three locations, than in the valid condition, in which the participant needed to monitor just one location. Hence, one may argue that perhaps our result can be explained by uncertainty reduction (Pelli, 1985; Tyler and Chen, 2000; Chen and Tyler, 2010). Our result did show a lower threshold in the valid condition than in the invalid conditions, and in turn a vertical shift of TvC functions that is consistent with uncertainty reduction. The three-fold increase in uncertainty from the valid to the invalid cued conditions, according to Tyler and Chen (2000), translated to a 2.5 dB threshold increment. This is slightly larger than the threshold difference between the valid and the invalid cue conditions in our data (2.2 dB). Furthermore, in our experiment, there were only two location-based cueing conditions (valid and invalid). The uncertainty effect, mathematically, as discussed in the Section “Model,” can be absorbed by a change of the sensitivity parameter,  $S_e$ . Thus, for practical reasons, we can consider the reduction of uncertainty as a cause of sensitivity change that accounts for the spatial cueing effect. However, uncertainty cannot explain the same-object advantage in our result. For instance, the TvC functions for the same-object and the different-object conditions were different even though the uncertainty in these two conditions was identical.

## CONCLUSION

The current study measured the thresholds in different levels of task difficulty and revealed the underlying mechanisms of location-based and object-based attention – which are difficult to evaluate from conventional RT measurements – and sheds a new light to current theories of object-based attention. Here, we overturn two widely accepted theories that object-based attention is due to the “spread” or “prioritization” of attention. In addition to revealing the underlying mechanisms of location- and object-based attention, the current finding fills the gap between previous physiological (Fink et al., 1997; He et al., 2004; Wager et al., 2004; He et al., 2008) and behavioral evidence (Shomstein and Yantis, 2004; List and Robertson, 2007; Chou and Yeh, 2008, 2011; Matsukura and Vecera, 2009) that have demonstrated the discrepancy in location-based and object-based attention by providing important convergent evidence from a novel aspect using the noise masking paradigm to the double-rectangle method.

## ACKNOWLEDGMENTS

This research was conducted as part of the first author's Ph.D. dissertation project (Chou, 2010). Preliminary versions were presented at the 2009 European Conference on Visual Perception

annual meeting (Chou et al., 2009). This study was supported by NSC-102-2410-H-431-008 and NSC-102-2420-H-431-001-MY2 to Wei-Lun Chou, by NSC101-2410-H-002-083-MY3 to Su-Ling Yeh, and by NSC 96-2413-H-002-006-MY3 to Chien-Chung Chen.

## REFERENCES

- Abrams, R. A., and Law, M. B. (2000). Object-based visual attention with endogenous orienting. *Percept. Psychophys.* 62, 818–833. doi: 10.3758/BF03206925
- Brown, J. M., Breitmeyer, B. G., Leighty, K. A., and Denney, H. I. (2006). The path of visual attention. *Acta Psychol.* 121, 199–209. doi: 10.1016/j.actpsy.2004.10.020
- Brawn, P. T., and Snowden, R. J. (2000). Attention to overlapping objects: detection and discrimination of luminance changes. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 342–358. doi: 10.1037/0096-1523.26.1.342
- Chen, C. C., and Foley, J. M. (2004). Pattern detection: interactions between oriented and concentric patterns. *Vision Res.* 44, 915–924. doi: 10.1016/j.visres.2003.11.017
- Chen, C. C., Foley, J. M., and Brainard, D. H. (2000). Detection of chromoluminance patterns on chromoluminance pedestals II: model. *Vision Res.* 40, 789–803. doi: 10.1016/S0042-6989(99)00228-X
- Chen, C. C., and Tyler, C. W. (1999). Spatial pattern summation is phase-insensitive in the fovea but not in the periphery. *Spat. Vis.* 12, 267–286. doi: 10.1163/156856899X00166
- Chen, C. C., and Tyler, C. W. (2001). Lateral sensitivity modulation explains the flanker effect in contrast discrimination. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 509–516. doi: 10.1098/rspb.2000.1387
- Chen, C. C., and Tyler, C. W. (2010). Symmetry: modeling the effects of masking noise, axial cueing and salience. *PLoS ONE* 5:e9840. doi: 10.1371/journal.pone.0009840
- Chou, W.-L. (2010). *Distinct Mechanisms of Location-Based and Object-Based Visual Attention*. Unpublished doctoral dissertation, National Taiwan University, Taiwan.
- Chou, W.-L., Chen, C. C., and Yeh, S.-L. (2009). Mechanisms of location- and object-based attention as revealed by the external noise paradigm. *Perception* 38(Suppl.), 1–194.
- Chou, W.-L., and Yeh, S.-L. (2008). Location- and object-based inhibition of return are affected by different kinds of working memory. *Quart. J. Exp. Psychol.* 61, 1761–1768. doi: 10.1080/17470210802194308
- Chou, W.-L., and Yeh, S.-L. (2011). Subliminal spatial cues capture attention and strengthen between-object link. *Conscious. Cogn.* 20, 1265–1271. doi: 10.1016/j.concog.2011.03.007
- Cohn, T. E., and Lasley, D. J. (1974). Detectability of a luminance increment: effect of spatial uncertainty. *J. Opt. Soc. Am.* 64, 1715–1719. doi: 10.1364/JOSA.64.001715
- Duncan, J. (1984). Selective attention and the organization of visual information. *J. Exp. Psychol. Gen.* 113, 501–517. doi: 10.1037/0096-3445.113.4.501
- Davis, G., and Driver, J. (1997). Spreading of visual attention to modally versus amodally completed regions. *Psychol. Sci.* 8, 275–281. doi: 10.1111/j.1467-9280.1997.tb00438.x
- Davis, G., Driver, J., Pavani, F., and Shepherd, A. (2000). Reappraising the apparent costs of attending to two separate visual objects. *Vision Res.* 40, 1323–1332. doi: 10.1016/S0042-6989(99)00189-3
- Doshier, B. A., and Lu, Z.-L. (2000). Mechanisms of perceptual attention in pre-cueing of location. *Vision Res.* 40, 1269–1292. doi: 10.1016/S0042-6989(00)00019-5
- Egely, R., Driver, J., and Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen.* 123, 161–176. doi: 10.1037/0096-3445.123.2.161
- Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of local attention: a zoom lens model. *Percept. Psychophys.* 40, 225–240. doi: 10.3758/BF03211502
- Eriksen, C. W., and Yeh, Y. Y. (1985). Allocation of attention in the visual field. *J. Exp. Psychol. Hum. Percept. Perform.* 11, 583–597. doi: 10.1037/0096-1523.11.5.583
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *J. Opt. Soc. Am.* 11, 1710–1719. doi: 10.1364/JOSAA.11.001710

- Foley, J. M., and Chen, C. C. (1999). Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: threshold measurements and a model. *Vision Res.* 39, 3855–3872. doi: 10.1016/S0042-6989(99)00104-2
- Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., and Frith, C. D. (1997). Space-based and object-based visual attention: shared and specific neural domains. *Brain* 120, 2013–2028. doi: 10.1093/brain/120.11.2013
- Gibson, B. S., and Egeth, H. (1994). Inhibition of return to object-based and environment-based locations. *Percept. Psychophys.* 55, 323–339. doi: 10.3758/BF03207603
- Goris, R. L., Wagemans, J., and Wichmann, F. A. (2008). Modeling contrast discrimination data suggest both the pedestal effect and stochastic resonance to be caused by the same mechanism. *J. Vis.* 8, 17.1–17.21. doi: 10.1167/8.15.17
- Han, S. M., Doshier, B. A., and Lu, Z.-L. (2003). Object attention revisited: boundary conditions and mechanisms. *Psychol. Sci.* 14, 598–604. doi: 10.1046/j.0956-7976.2003.psci\_1471.x
- He, X., Fan, S., Zhou, K., and Chen, L. (2004). Cue validity and object-based attention. *J. Cogn. Neurosci.* 16, 1085–1097. doi: 10.1162/0898929041502689
- He, X., Humphreys, G., Fan, S., Chen, L., and Han, S. (2008). Differentiating spatial and object-based effects on attention: an event-related brain potential study with peripheral cueing. *Brain Res.* 1245, 116–125. doi: 10.1016/j.brainres.2008.09.092
- Kasai, T., and Kondo, M. (1997). Electrophysiological correlates of attention-spreading in visual grouping. *Neuroreport* 18, 93–98. doi: 10.1097/WNR.0b013e328011b8c9
- Kontsevich, L. L., and Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res.* 39, 2729–2737. doi: 10.1016/S0042-6989(98)00285-5
- LaBerge, D., and Brown, V. (1989). Theory of attentional operations in shape identification. *Psychol. Rev.* 96, 101–124. doi: 10.1037/0033-295X.96.1.101
- Lu, Z.-L., and Doshier, B. A. (1998). External noise distinguishes mechanisms of attention. *Vis. Res.* 38, 1183–1198. doi: 10.1016/S0042-6989(97)00273-3
- Lu, Z.-L., and Doshier, B. A. (2000). Spatial attention: different mechanisms for central and peripheral temporal precues? *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1534–1548. doi: 10.1037/0096-1523.26.5.1534
- Liu, S.-H., Doshier, B. A., and Lu, Z.-L. (2009). The role of judgment frames and task precision in object attention: reduced template sharpness limits dual-object performance. *Vision Res.* 49, 1336–1351. doi: 10.1016/j.visres.2008.07.025
- Legge, G. E., Kersten, D., and Burgess, A. E. (1987). Contrast discrimination in noise. *J. Opt. Soc. Am.* 4, 391–404. doi: 10.1364/JOSA.4.000391
- List, A., and Robertson, L. C. (2007). Inhibition of return and object-based attentional selection. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1322–1334. doi: 10.1037/0096-1523.33.6.1322
- Lamy, D., and Tsai, Y. (2000). Object features, object locations, and object files: which does selective attention activate and when? *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1387–1400. doi: 10.1037/0096-1523.26.4.1387
- Moore, C. M., and Fulton, C. (2005). The spread of attention to hidden portions of occluded surfaces. *Psychon. Bull. Rev.* 12, 301–306. doi: 10.3758/BF03196376
- Matsukura, M., and Vecera, S. P. (2006). The return of object-based attention: selection of multiple-region objects. *Percept. Psychophys.* 68, 1163–1175. doi: 10.3758/BF03193718
- Matsukura, M., and Vecera, S. P. (2009). Interference between object-based attention and object-based memory. *Psychon. Bull. Rev.* 16, 529–536. doi: 10.3758/PBR.16.3.529
- Moore, C. M., Yantis, S., and Vaughan, B. (1998). Object-based visual selection: evidence from perceptual completion. *Psychol. Sci.* 9, 104–110. doi: 10.1111/1467-9280.00019
- Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *J. Opt. Soc. Am.* 54, 950–955. doi: 10.1364/JOSA.54.000950
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am.* 2, 1508–1532. doi: 10.1364/JOSA.2.001508
- Pelli, D. G. (1991). “Noise in the visual system may be early,” in *Computational Models of Visual Processing*, eds M. Landy and J. A. Movshon (Cambridge: MIT Press), 147–152.
- Pestilli, F., and Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Res.* 45, 1867–1875. doi: 10.1016/j.visres.2005.01.019
- Posner, M. I. (1980). Orienting of attention. *Quart. J. Exp. Psychol.* 32, 3–25. doi: 10.1080/00335558008248231
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108. doi: 10.1037/0033-295X.85.2.59
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002
- Richard, A. M., Lee, H., and Vecera, S. P. (2008). Attentional spreading in object-based attention. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 842–852. doi: 10.1037/0096-1523.34.4.842
- Ross, J., and Speed, H. D. (1991). Contrast adaptation and contrast masking in human vision. *Proc. Biol. Sci.* 246, 61–70. doi: 10.1098/rspb.1991.0125
- Shomstein, S., and Behrmann, M. (2008). Object-based attention: strength of object representation and attentional guidance. *Percept. Psychophys.* 70, 132–144. doi: 10.3758/PP.70.1.132
- Shomstein, S., and Yantis, S. (2002). Object-based attention: sensory modulation or priority setting? *Percept. Psychophys.* 64, 41–51. doi: 10.3758/BF03194556
- Shomstein, S., and Yantis, S. (2004). Configural and contextual prioritization in object-based attention. *Psychon. Bull. Rev.* 11, 247–253. doi: 10.3758/BF03196566
- Snowden, R. J., and Hammett, S. T. (1998). The effects of surround contrast on contrast thresholds, perceived contrast and contrast discrimination. *Vision Res.* 38, 1935–1945. doi: 10.1016/S0042-6989(97)00379-9
- Tipper, S. P., Driver, J., and Weaver, B. (1991). Object-centred inhibition of return of visual attention. *Quart. J. Exp. Psychol.* 43A, 289–298. doi: 10.1080/14640749108400971
- Teo, P. C., and Heeger, D. K. (1994). Perceptual image distortion. *SPIE Proc.* 2179, 127–141. doi: 10.1117/12.172664
- Tyler, C. W., and Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: attention, channel uncertainty and probability summation. *Vision Res.* 40, 3121–3144. doi: 10.1016/S0042-6989(00)00157-7
- Tyler, C. W., and McBride, B. (1997). The Morphonome image psychophysics software and a calibrator for Macintosh systems. *Spat. Vis.* 10, 479. doi: 10.1163/156856897X00410
- Wager, T. D., Jonides, J., and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* 22, 1679–1693. doi: 10.1016/j.neuroimage.2004.03.052
- Watson, A. B., and Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am.* 4, 2379–2391. doi: 10.1364/JOSA.4.002379
- Wilson, H. R., and Humanski, R. (1993). Spatial frequency adaptation and contrast gain control. *Vision Res.* 33, 1133–1149. doi: 10.1016/0042-6989(93)90248-U
- Wu, C. C., and Chen, C. C. (2010). Distinguishing lateral interaction from uncertainty reduction in collinear flanker effect on contrast discrimination. *J. Vis.* 10, 1–14. doi: 10.1167/10.3.8
- Zenger, B., Braun, J., and Koch, C. (2000). Attentional effects on contrast detection in the presence of surround masks. *Vision Res.* 40, 3717–3724. doi: 10.1016/S0042-6989(00)00218-2

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 January 2014; accepted: 28 April 2014; published online: 21 May 2014.

Citation: Chou W-L, Yeh S-L and Chen C-C (2014) Distinct mechanisms subserve location- and object-based visual attention. *Front. Psychol.* 5:456. doi: 10.3389/fpsyg.2014.00456

This article was submitted to *Perception Science*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Chou, Yeh and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Evidence for adjustable bandwidth orientation channels

Christopher P. Taylor<sup>1\*</sup>, Patrick J. Bennett<sup>2,3\*</sup> and Allison B. Sekuler<sup>2,3</sup>

<sup>1</sup> Department of Psychology and Clinical Language Sciences, Centre for Integrative Neuroscience and Neurodynamics, University of Reading, Reading, UK

<sup>2</sup> Department of Psychology, Neuroscience, and Behaviour, McMaster University, Hamilton, ON, Canada

<sup>3</sup> Centre for Vision Research, York University, Toronto, ON, Canada

## Edited by:

Denis Pelli, New York University,  
USA

## Reviewed by:

John Cass, University of Western  
Sydney, Australia

Daniel Hart Baker, University of York,  
UK

Keith Anthony May, UCL, UK

## \*Correspondence:

Christopher P. Taylor, Department of  
Psychology and Clinical Language  
Sciences, Centre for Integrative  
Neuroscience and Neurodynamics,  
University of Reading, Whiteknights,  
Shinfield Rd, Reading,  
West Berkshire RG6 6AL, UK  
e-mail: christopher.taylor@  
gmail.com;

Patrick J. Bennett, Department of  
Psychology, Neuroscience, and  
Behaviour, McMaster University,  
1280 Main Street West, Hamilton,  
ON L8S 4K1, Canada  
e-mail: bennett@mcmaster.ca

The standard model of early vision claims that orientation and spatial frequency are encoded with multiple, quasi-independent channels that have fixed spatial frequency and orientation bandwidths. The standard model was developed using detection and discrimination data collected from experiments that used deterministic patterns such as Gabor patches and gratings used as stimuli. However, detection data from experiments using noise as a stimulus suggests that the visual system may use adjustable-bandwidth, rather than fixed-bandwidth, channels. In our previous work, we used classification images as a key piece of evidence against the hypothesis that pattern detection is based on the responses of channels with an adjustable spatial frequency bandwidth. Here we tested the hypothesis that channels with adjustable orientation bandwidths are used to detect two-dimensional, filtered noise targets that varied in orientation bandwidth and were presented in white noise. Consistent with our previous work that examined spatial frequency bandwidth, we found that detection thresholds were consistent with the hypothesis that observers sum information across a broad range of orientations nearly optimally: absolute efficiency for stimulus detection was 20–30% and approximately constant across a wide range of orientation bandwidths. Unlike what we found with spatial frequency bandwidth, the results of our classification image experiment were consistent with the hypothesis that the orientation bandwidth of internal filters were adjustable. Thus, for orientation summation, both detection thresholds and classification images support the adjustable channels hypothesis. Classification images also revealed hallmarks of inhibition or suppression from uninformative spatial frequencies and/or orientations. This work highlights the limitations of the standard model of summation for orientation. The standard model of orientation summation and tuning was chiefly developed with narrow-band stimuli that were not presented in noise, stimuli that are arguably less naturalistic than the variable bandwidth stimuli presented in noise used in our experiments. Finally, the disagreement between the results from our experiments on spatial frequency summation with the data presented in this paper suggests that orientation may be encoded more flexibly than spatial frequency channels.

**Keywords:** pattern vision, orientation, channels, ideal observer, classification images, summation, psychophysics

## 1. INTRODUCTION

Visual noise has been used to investigate visual processing in a variety of tasks (Pelli and Farell, 1999). Researchers have used two-dimensional luminance noise most frequently, but studies have also used one-dimensional luminance noise (i.e., noise that is constrained to vary along a single dimension) as well as visual noise that varies in other ways such as color (Gegenfurtner and Kiper, 1992), motion (Dakin et al., 2005), orientation (Girshick et al., 2011), Gaussian spatial windowing or bubbles (Gosselin and Schyns, 2001), and zero-dimensional noise pedestal increments (Baker and Meese, 2012). In virtually all of these studies, the noise was used as a mask and the observer's task was to ignore the noise to detect a non-noise target. Comparatively few studies have used noise as the target stimulus itself. David Green and colleagues used noise in this way to study the mechanisms underlying the detection of auditory signals (Green, 1960a,b; Green and

Swets, 1966), and subsequent studies adopted Green's approach to study vision (for examples, see Kersten, 1987; Taylor et al., 2003, 2004, 2005, 2006, 2009; Levi et al., 2005, 2008). In the current study, we use noise targets and noise masks to investigate orientation selectivity of visual mechanisms.

Data from detection, discrimination, and adaptation experiments using both psychophysical and physiological methods support the idea that the early stages of visual processing encode patterns with channels that are tuned to a fixed range of spatial frequency and orientation (Campbell and Kulikowski, 1966; Campbell et al., 1966; Graham, 1989; Wandell, 1995). This standard, or back-pocket, model of early visual coding accounts for a wide range of detection and discrimination data (Klein, 1992; Watson, 2000). However, despite its many successes, the standard multiple channels model apparently fails to account for some experimental results (see Nachmias et al., 1973; Kersten,

1987; Derrington and Henning, 1989; Perkins and Landy, 1991; Wandell, 1995; Taylor et al., 2009). One particularly puzzling result was reported by Kersten (1987), who measured detection thresholds for a visual noise target embedded in a visual noise mask. Detection thresholds were measured with noise targets with different spatial frequency bandwidths. For frequency bandwidths between 1 and 4 octaves, Kersten found that detection thresholds were proportional to the quarter-root of bandwidth,

$$c_{rms} \propto BW^{\frac{1}{4}} \quad (1)$$

Interestingly, Kersten showed that detection thresholds for an ideal observer also were proportional to the quarter-root of bandwidth, which implies that absolute efficiency ( $\eta$ ), defined as

$$\eta = \left( \frac{c_{rms} (ideal)}{c_{rms} (observer)} \right)^2 \quad (2)$$

ought to be constant. Indeed, Kersten found that absolute efficiency was high ( $\approx 50\%$ ) and approximately constant as the spatial frequency bandwidth of the noise stimulus was increased from 0.5 to 4 octaves. Kersten pointed out that this result is surprising because ideal observers integrate information across the entire spatial frequency bandwidth, whereas human observers are thought to detect patterns using mechanisms that have bandwidths that are much narrower than four octaves. Hence, the data suggest that spatial frequency summation is approximately optimal across a wide bandwidth, and appear to be inconsistent with a standard model that assumes that patterns are detected using channels that have a fixed and relatively narrow frequency bandwidth. Instead, Kersten suggested that the data were consistent with the adjustable channels hypothesis, first proposed by Green (1960a,b) to explain similar results obtained in an auditory detection task, which states that human observers detect band-limited noise using a channel, or combination of channels, with a frequency bandwidth that is adjusted to match that of the stimulus, and which sums information efficiently across the entire bandwidth.

Taylor et al. (2009) evaluated the adjustable channels hypothesis by using the classification image technique (Murray, 2011) to measure the frequencies observers use to detect visual noise that varied in bandwidth from 0.5 to 6 octaves. Like Kersten (1987), Taylor et al. found that detection thresholds were proportional to the quarter-root of bandwidth (Equation 1). However, contrary to the predictions of the adjustable channels hypothesis, estimates of the spatial frequency bandwidth of the channel used to detect visual noise, which was derived from the classification image data, did not vary with stimulus bandwidth. Furthermore, Taylor et al. used Monte Carlo simulations to demonstrate that the optimal spatial frequency summation found in noise detection tasks was, surprisingly, consistent with the predictions of at least one version of the standard model (Wilson et al., 1983). In short, Taylor et al. showed that the apparently anomalous results reported by Kersten (1987) were consistent with standard models of spatial frequency summation (Graham, 1989).

In this paper, we follow up on our previous work on spatial frequency summation and investigate orientation summation in two

experiments. The first experiment measures detection thresholds and absolute efficiency of noise patterns that vary in orientation bandwidth. The second experiment uses the classification image technique to estimate the tuning characteristics of the internal filters used in this noise detection task. To anticipate our results, we find that orientation summation is, like spatial frequency summation, nearly optimal across a wide range of bandwidths. However, unlike what was found with frequency summation, the classification image results are consistent with the hypothesis that the orientation bandwidth of the internal filter that mediates detection is adjusted to match the stimulus.

## 2. EXPERIMENT 1

### 2.1. MATERIALS AND METHODS

#### 2.1.1. Observers

The three observers (all female; 23–27 years of age) in this experiment were members of the McMaster University community and were paid for their participation. Informed consent was obtained from all participants and the research approved by the McMaster University research ethics board. All observers were naïve about the experimental hypotheses, had normal or corrected-to-normal Snellen acuity and Pelli–Robson contrast sensitivity, and had extensive practice with this and other visual psychophysical tasks.

#### 2.1.2. Apparatus

Stimuli were generated and displayed using an Apple Macintosh G4 computer with an ATI Radeon video card running MATLAB and the Psychophysics and Video toolboxes (Brainard, 1997; Pelli, 1997). The stimulus display was a Sony GDM-F520 monitor set to a resolution of  $1024 \times 768$  pixels and subtended a visual angle of  $10.8^\circ \times 8.3^\circ$  at the viewing distance of 2 m. The frame rate of the display was 75 Hz and the mean luminance  $45 \text{ cd/m}^2$ . Display luminance was calibrated using a PhotoResearch PR-650 photometer before each session. The results of the calibration were used to linearize the display for that session, but in general there was little variability in the display from session to session. A Cambridge Research System Bits++ device was used to achieve fine grained (i.e., 14-bit) control of contrast. A custom designed button box with an ActiveWire card was used to record the observer's responses.

#### 2.1.3. Stimuli

The stimuli were two-dimensional Gaussian white noise patterns that were spatially filtered digitally with ideal (hard-edged) spatial frequency and orientation filters. The filter had a fixed center spatial frequency of 5 cy/deg, but depending on the experimental condition the spatial frequency bandwidth was either one or two octaves. The center orientation of the filter was horizontal. The two-sided orientation bandwidths were  $2^\circ$ ,  $8^\circ$ ,  $16^\circ$ ,  $32^\circ$ ,  $64^\circ$ ,  $128^\circ$ , and  $180^\circ$ . For example, a two-sided orientation bandwidth filter of  $16^\circ$  passed orientations from  $-8^\circ$  to  $+8^\circ$ . To prevent edge artifacts, stimulus contrast was modulated on the screen with a circularly-symmetric Gaussian envelope with a standard deviation of  $1.08^\circ$  of visual angle. A white noise mask with a contrast variance of 0.32 was used in all conditions to mask the signal noise. On each trial, a new sample of signal noise and background noise were generated on each interval on



every trial. The monitor provided the only illumination in the testing room.

### 2.1.4. Procedure

Observers viewed the stimuli binocularly through natural pupils. A two-interval forced-choice (2-IFC) procedure was used. The observer was instructed to fixate a high-contrast dot located in the center of the display. The observer initiated each trial by pressing the space-bar on the keyboard. After a delay of 50 ms, the fixation point was removed, then after another 50 ms delay the first stimulus interval appeared. The first stimulus interval was 200 ms in duration and was followed by a 300 ms blank inter-stimulus interval and then a second 200 ms stimulus interval. The two stimulus intervals were marked by clearly audible tones, and a high/low pitched tone indicated whether a response was correct/incorrect. The observer's task was to determine which of the two stimulus intervals contained the target.

Stimulus contrast variance was varied across trials using four interleaved staircases, two converging on the 71% correct point of the psychometric function and two on the 84% correct point (Wetherill and Levitt, 1965). The staircases were stopped when the observer had completed 75 trials in each staircase. The total number of trials in each session was 2100 (300 trials per stimulus bandwidth, and seven stimulus bandwidths/session). Thresholds, defined as the RMS contrast required to produce 75% correct, were estimated by fitting a cumulative normal to all the data collected.

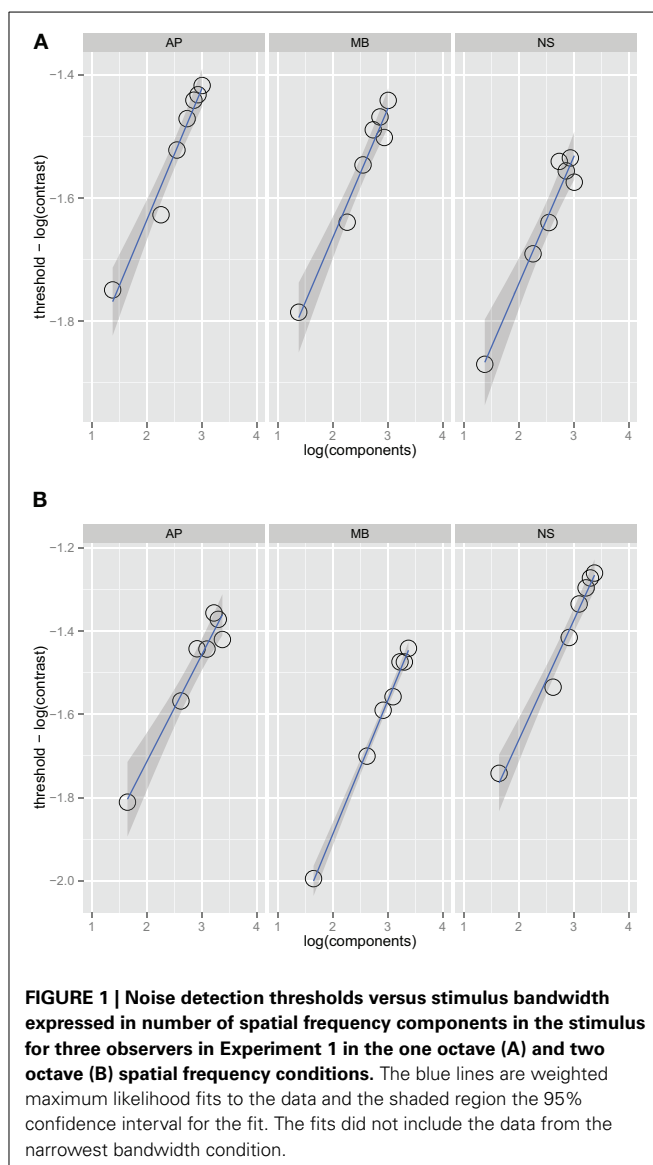
A  $3^\circ \times 3^\circ$  square, drawn with a high-contrast, 2-pixel wide line, was centered on the fixation point surrounded the stimulus to reduce spatial uncertainty. The frame remained on the screen for the entire duration of each trial: it was centered on the fixation point at the start of a trial, and remained visible until the observer made a response. To reduce adaptation, the square had a 50% probability of being black or white on each trial.

Thresholds were measured with stimuli that had spatial frequency bandwidths of 1 or 2 octaves. Two spatial frequency bandwidth conditions were run in separate sessions, alternating with each session, and each observer began with a spatial frequency bandwidth chosen randomly. In each test session, orientation bandwidths were presented in separate blocks of trials and the order of bandwidths was randomized. All orientation bandwidth conditions were completed during a single session.

In addition to the conditions described above, we measured contrast detection thresholds for a white noise stimulus as control in a separate session for the same observers. The three observers completed four white noise detection sessions, each containing 300 trials. The thresholds from this control condition were collected after all other conditions in Experiment 1 were completed.

## 2.2. RESULTS

**Figure 1** shows threshold versus bandwidth (TvB) functions for three observers in the one and two octave spatial frequency bandwidth conditions. We checked for interval effects to determine if there was a difference in threshold for stimuli presented in the first and second interval (Yeshurun et al., 2008), but there were no significant threshold differences between the two intervals.



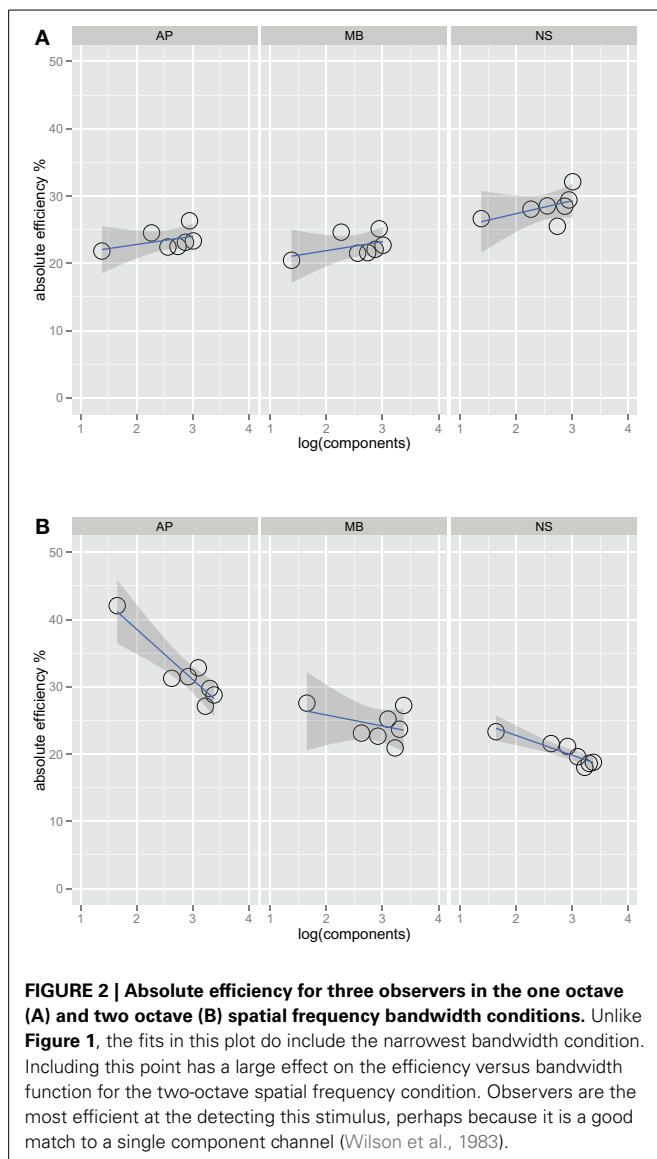
**FIGURE 1 | Noise detection thresholds versus stimulus bandwidth expressed in number of spatial frequency components in the stimulus for three observers in Experiment 1 in the one octave (A) and two octave (B) spatial frequency conditions.** The blue lines are weighted maximum likelihood fits to the data and the shaded region the 95% confidence interval for the fit. The fits did not include the data from the narrowest bandwidth condition.

Each facet of the figure shows the thresholds for an observer (AP, MB, and NS) versus orientation bandwidth. Two-sided orientation bandwidth, which varied from  $2^\circ$  to  $180^\circ$ , is expressed as the number of Fourier components in the stimulus because the ideal observer's threshold depends on the number of components, rather than the orientation bandwidth *per se*. **Figure 1** shows that detection threshold, when expressed as the logarithm of RMS contrast, increases with increasing orientation bandwidth. Each point on the graph corresponds to one of the orientation bandwidth conditions, the left-most to  $2^\circ$  and the right-most to  $180^\circ$ . There are no statistical differences between the thresholds in the one and two octave spatial frequency bandwidth conditions. The narrowest bandwidth condition was not included in the fitting procedure. If stimuli are sufficiently narrow-band then the TvB function will flatten out, producing what has been referred to in the literature as the critical-band (Quick et al., 1976). Characterizing the critical band was not the focus of this work and thus, we excluded the narrowest bandwidth, but if this



point was included in the analysis, the TvB function would flatten out and have a shallower slope. Thresholds in all observers and conditions were well fitted by a power function. Bootstrap confidence intervals were calculated by simulating 999 fits to the observer data—the 95% confidence intervals always included 0.25 and ranged from 0.23 to 0.26. Finding a slope 0.25 in is line with the prediction of the quarter-root law and indicates optimal summation.

**Figures 2A,B** show absolute efficiency (Equation 2) as a function of orientation bandwidth for the two spatial frequency conditions. For spatial frequency summation, absolute efficiency as high as 50% have been found (Taylor et al., 2009); in this experiment, absolute efficiency was also relatively high, ranging from 20% to 40%. Thresholds for the ideal observer were computed via simulations that were approximations to a two-dimensional version of that found in previous work (Kersten, 1987; Taylor et al., 2009).

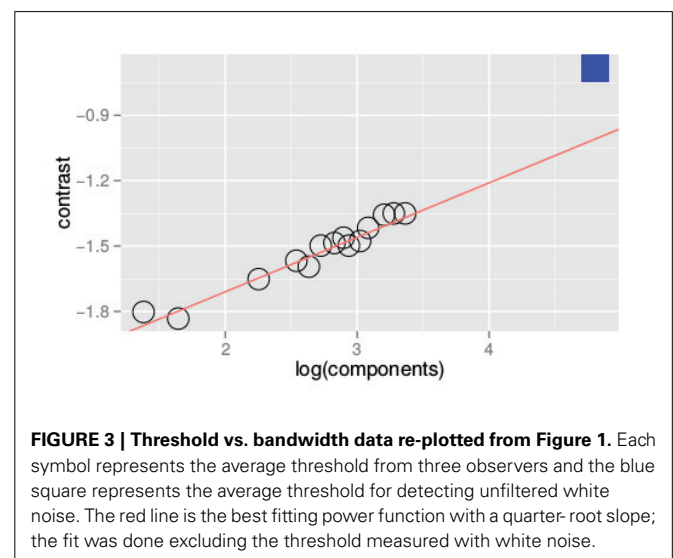


**Figure 3** is a summary figure of the data in **Figure 1** and shows average detection thresholds in each spatial frequency and bandwidth condition plotted against the number of frequency components in the stimulus. The red line in the figure depicts the prediction of optimal summation for the quarter-root law. The blue square represents the mean white noise threshold, expressed in RMS contrast, for the three observers. The white noise threshold data point was not used when the combined data were fit. Including this data point provides an instructive test as it demonstrates that the quarter-root law breaks down when the stimulus includes all frequencies and orientations. Although the quarter-root law breaks down for white noise, the number of components required to observe a breakdown of the quarter-root law has yet to be determined. White noise thresholds suggest that if there is channel adjustment, there are limits to the adjustment that remain to be characterized.

The results of Experiment 1 are similar to the results for spatial frequency summation (Kersten, 1987; Taylor et al., 2009) and auditory noise detection (Green, 1960a,b) in that the TvB functions have a quarter-root slope, the same slope produced by an ideal observer. Quarter-root TvB slopes, along with the high absolute efficiencies we observed, are consistent with the idea that orientation information is summed optimally. Both of these findings are necessary but not sufficient to conclude that noise is detected by adjustable channels. As shown by our previous work on spatial-frequency summation (Taylor et al., 2009), it is important to pair estimates of threshold with the classification image method to characterize the channel used by observers. Classification images can change the interpretation of the TvB function substantially; for spatial frequency classification images lead us to interpret our data as supporting a fixed-channel model rather than an adjustable-channel model.

### 3. EXPERIMENT 2

In Experiment 2, we measured classification images with orientation filtered noise in a sub-set of conditions used in Experiment 1.



### 3.1. MATERIALS AND METHODS

#### 3.1.1. Observers

The two observers were 28-year old students at McMaster University who were paid for their participation. Both observers were unaware of the experimental hypotheses, had normal Snellen acuity, had extensive practice in psychophysical tasks, and participated in Experiment 1.

#### 3.1.2. Apparatus

The apparatus was identical to that used in Experiment 1.

#### 3.1.3. Stimuli

The stimuli and noise had the same parameters as those used in the one octave spatial frequency bandwidth condition in Experiment 1.

#### 3.1.4. Procedures

The procedure of Experiment 2 was the same as Experiment 1 except that the contrast of the stimulus was held constant at the 75% threshold measured in Experiment 1. There were 2500 trials per condition or classification image, for a total of 7500 trials per observer.

### 3.2. RESULTS

We measured our classification images using a two-interval forced choice method rather than a yes/no procedure as described by Abbey et al. (1999) and calculated our classification images using the *power spectra* of the noise masks, rather than the noise masks themselves. This produces classification images in the power spectrum which has been used previously by Solomon (2002). Because the method is described in detail elsewhere, only a brief description is provided here. On each trial, the power spectrum of the noise mask in each interval was computed, the difference between the pair of power spectra calculated, and finally the difference spectrum was placed into one of four bins based on which interval contained the signal (1 or 2) and the observer's response (correct or incorrect). The difference power spectra were then averaged by the number of trials in that bin and then the two average spectra computed from correct responses were averaged, as were the two average spectra computed from incorrect trials. Finally, the difference between the correct and incorrect averaged spectra was computed and the resulting classification image was normalized to have a peak value of one. Classification images calculated using this procedure are proportional to the linear template applied to the power spectra (Abbey et al., 1999; Abbey and Eckstein, 2002).

**Figure 4** shows the raw classification images for the ideal observer and two human observers. Each classification image was computed using the same number of trials. The images represent spatial frequency as the distance from the center of the image. Orientation information is represented by sets of pixels in a line that begins in the center of the image and extends to its edge. The power spectra have been rotated so that the horizontal and vertical orientations in the stimulus correspond to the central horizontal row and vertical column of pixels in the image. The gray level of each pixel in the classification image represents how the power of an individual Fourier component is weighted by the

observer when performing the noise detection task. If the pixel is lighter than median gray, then noise power at that frequency and orientation is positively correlated with the probability of a correct response; the lighter the pixel, the higher the correlation. Conversely, for pixels darker than median gray, power at that frequency and orientation is negatively correlated with the probability of a correct response.

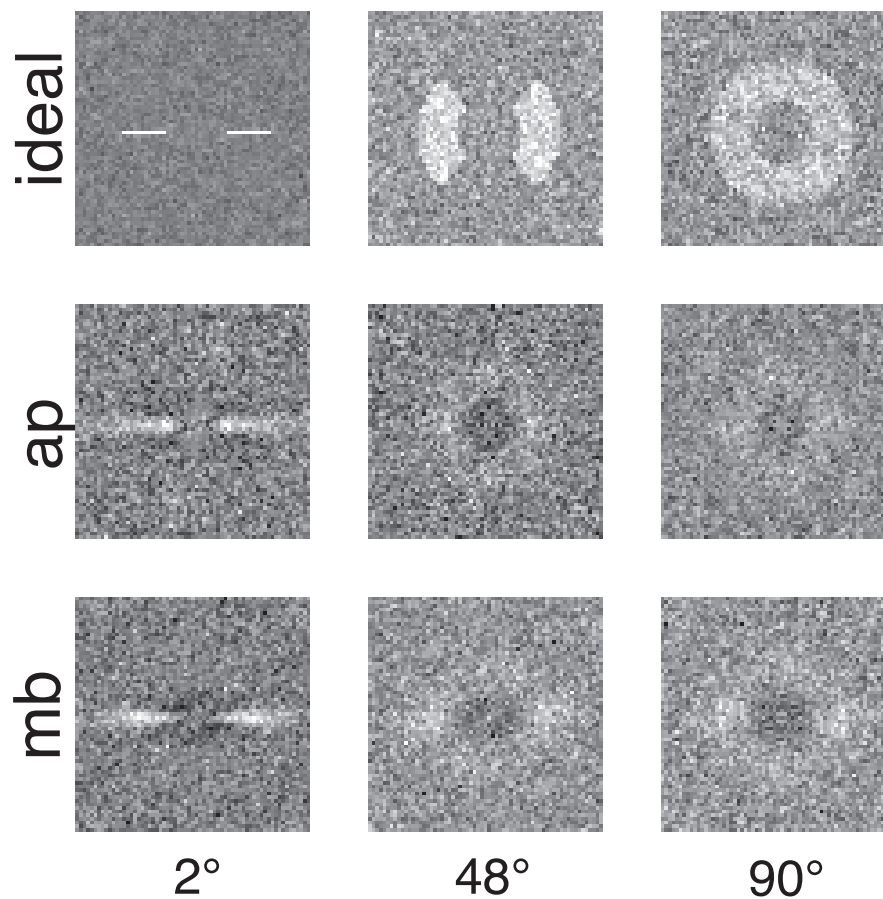
The classification images shown in **Figure 4** are  $64 \times 64$  subsets of the full  $512 \times 512$  power spectra which correspond to spatial frequencies from DC to approximately 20 cy/deg and include the spatial frequencies presented in the stimulus. **Figure 5** shows classification images that have been smoothed with a  $5 \times 5$  triangular convolution kernel (equivalent to linear interpolation) to reduce spurious noise in the template that results from a limited number of trials.

**Figures 4, 5** show several important results. First, the human observers' classification images resemble those of the ideal observer, in that they have a narrow bandwidth (as measured by half-width at half-height) with the smallest stimulus bandwidth and get larger with increasing stimulus. Bandwidths of the classification images in the  $48^\circ$  and  $90^\circ$  conditions were larger than the bandwidths measured in the  $2^\circ$  condition. Also, the classification images from human observers have pronounced dark regions at off-stimulus orientations and frequencies that are not present in the classification images for the ideal observer. Noise power at these Fourier components was negatively correlated with the probability of correctly detecting the signal, an important finding that will be returned to in the discussion.

### 3.3. ANALYSIS

To relate the classification images to orientation channels found in orientation masking experiments (e.g., Govenlock et al., 2009), the two-dimensional classification images collected in this experiment were collapsed into one-dimensional classification images as a function of orientation. Values in each classification image were summed in  $1^\circ$  steps across a band spatial frequencies (filter center-frequency 5 cy/deg and bandwidth of approximately 20 cy/deg) over a  $180^\circ$  range of orientations. The resulting values are plotted in **Figure 6**. Two features of the data are readily apparent. First, orientations around  $0^\circ$  (i.e., horizontal) had the strongest influence on observers' decisions. Second, vertical orientations or other orientations far away from zero had a weaker influence on decisions that was opposite to that of horizontal frequencies.

We fit a Difference of Gaussians (DoG) function to our circularly summed normalized classification images. Classification images were normalized to the peak response. We chose DoG functions because preliminary analyses indicated that they fitted the data better than a single Gaussian and because DoG functions have been used previously to model orientation channels (De Valois et al., 1982; DeValois and DeValois, 1988; Carandini and Ringach, 1997; Ringach, 1998). We fixed the relative amplitude of the excitatory Gaussian to be twice that of the inhibitory Gaussian which is consistent with previous physiology (Sceniak et al., 2001). A DoG function has four free shape parameters—one for the center/mean and another for the bandwidth/standard deviation for each of the positive and negative Gaussians that



**FIGURE 4 | Classification images for the ideal observer and two human observers.** Low spatial frequencies are represented by the pixels in the center of the image and frequency increases with increasing distance from the center of the image. Horizontal orientations are represented in the center row of the image and

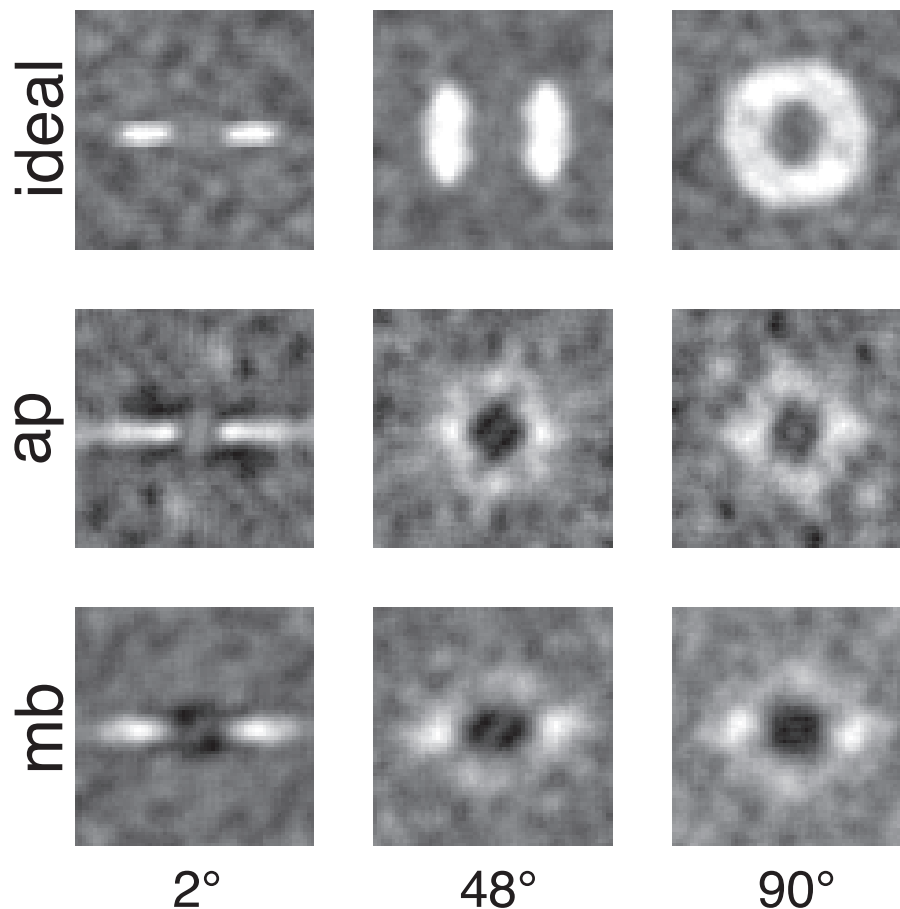
vertical orientations by the center column. All images were collected with the same number of trials. The ideal observer's classification image also serves as a depiction of filters used to generate the stimuli because the ideal observer only has access to the information in the filtered noise stimulus.

comprise the function—but we applied two constraints that were consistent with previous models of orientation channels (Burr et al., 1981; Ringach, 1998; Shirazi, 2004). The constraints were (1) both Gaussian functions were fixed to a common center; and (2) the bandwidth of the positive Gaussian was set to be narrower than the bandwidth of the negative Gaussian. **Figure 7** shows the best-fitting (least-squares) parameters and 95% confidence intervals computed via a percentile bootstrap procedure (Efron and Tibshirani, 1994). The center orientation of the best-fitting function did not change as the bandwidth of the stimulus was increased and was not different from zero, or horizontal (i.e., the orientation of the signal). The linear increase in the bandwidth for the negative Gaussian component (60–90°) was larger than the linear increase for the positive Gaussian (20–30°), although the proportional increase was about the same (i.e., 50%). Inhibitory mechanisms may be more flexible/adjustable in their responses than excitatory mechanisms. This hypothesis is supported by the data in **Figure 7**, specifically that slope of the red line is larger than that of the blue line.

#### 4. DISCUSSION

Our classification images support the adjustable channels hypothesis, unlike what was found in spatial frequency summation experiments (Taylor et al., 2009) using similar methods. This result implies that, contrary to the assumptions of the standard model, the mechanisms that produce optimal spatial frequency and orientation summation differ.

In the data, this point is illustrated by the negative weights in the classification images for orientation summation that occur at low spatial frequencies at all orientations but positive weights at higher spatial frequencies at a range of orientations dependent upon the signal (see **Figure 5**). The 1D templates derived from the 2D classification images also exhibit regions of suppression at orientations far removed from the center stimulus orientation (**Figure 6**), which correspond to the black/dark regions in the 2D classification images. Ideal templates do not show regions of suppression, thus the negative weights must be the result of psychological process. Furthermore, these negative weights were not found in spatial frequency summation experiments (Taylor et al.,



**FIGURE 5 | Smoothed classification images for the same conditions and observers presented in Figure 4.** White regions represented where power in the stimulus is correlated with a correct detection response, whereas the dark or black regions indicate regions where

incorrect responses are correlated with the stimulus. An important comparison to note is the difference between human and ideal observers at off-frequency and off-orientations in the stimulus. See the text for more detail.

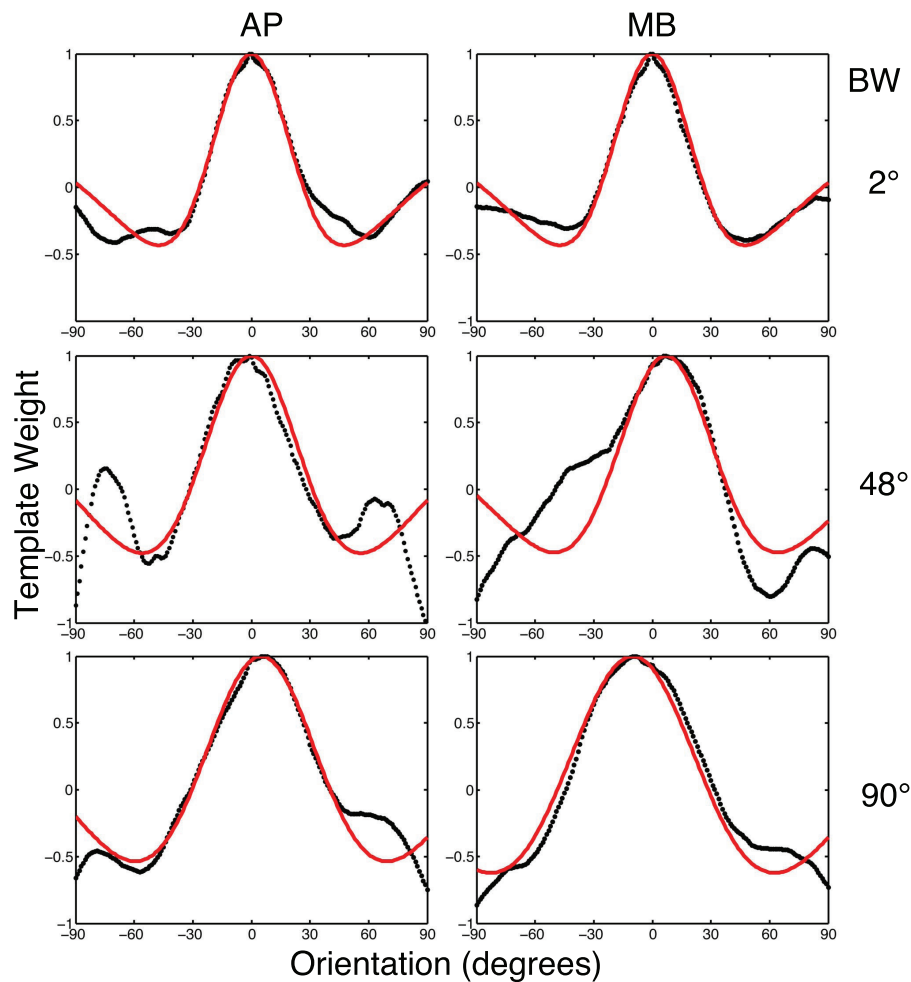
2009), and therefore appear to be specific for orientation summation. One interpretation of these dark bands is that they reflect the contribution of inhibitory orientation processing found both psychophysically and physiologically (Ringach, 1998; Ringach et al., 2002).

The orientation bandwidths of 1D templates measured by the classification image technique become broader with increasing stimulus bandwidth (see Figure 5), but the increase in bandwidth is smaller than the channel adjustment predicted by the ideal observer. The adjustability of human observers detection mechanisms is constrained by some, as yet unknown, process. Perhaps more complex, non-linear, biologically inspired modeling (Goris et al., 2013) can capture our results, but this remains to be tested.

A possible explanation for the differences between human and ideal templates is that human observers perform the detection task by differencing the power of different spatial frequency and/or orientation components. The ideal observer knows the center spatial frequency and orientation exactly; it also knows the spatial frequency and orientation bandwidth exactly. Human observers, may not have precise access to these four signal

parameters, even after many thousands of detection trials. Thus, human observers turn to an alternative strategy, one which we'll call a differencing strategy.

In Figure 5 one can see that observers use non-informative regions of the signal—anywhere the human classification image differs from the ideal classification image, this is the hallmark of the use of non-informative information (i.e., noise). Despite using non-informative information, human efficiency is still relatively high in the current task compared to efficiency in many other visual tasks (Gold et al., 1999). Why do observers use non-informative information? One hypothesis is that observers need to anchor their detection judgements and then compute a difference based on this perceptual anchor. According to this idea, observers can only make a detection decision based on the relative power within two (or perhaps more) regions of the power spectrum. In our task, the light and dark regions may represent the portions of the power spectrum that are being compared: observers may be basing their decisions on the difference between power at low spatial frequencies (at all orientations) and power at spatial frequencies and orientations within the signal band. In



**FIGURE 6 |** The black dotted lines show the one-dimensional circularly summed classification images for two observers in three conditions. The red lines show the best difference of Gaussian fits to the empirical data. The x-axis in each sub-figure is orientation, the y-axis the weight of the classification image.

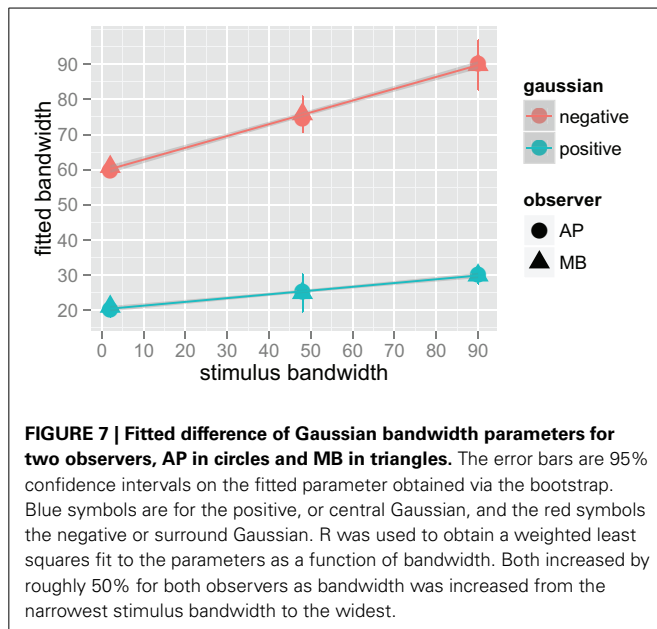
a given interval, if the power in the signal band is high and the power within the inhibitory region is low, observers will select that interval as the one that contains the signal. If however, the power in the inhibitory region is high and the in the signal band low, the observers will actively choose to not select that stimulus interval as containing the signal.

What is the functional role of the measured inhibitory mechanisms? One possibility, backed up by a great deal of evidence is that they play a role in contrast gain control (e.g., Watson and Solomon, 1997; Schwartz and Simoncelli, 2001). An alternative idea is that the visual system contains mechanisms that signal whether a stimulus ought to be considered an edge or a part of a texture. This hypothesis is inspired by the work on “end-stopping” found in the motion (Pack et al., 2003) and contour (Heitger et al., 1992) literature. The inhibitory mechanisms revealed by our classification images might provide a sort of end-stopping in Fourier space that limits the information that is combined into an edge or a texture. To be specific, if the orientation bandwidth within a region of visual space, as signalled

by suppressive mechanisms is narrow, then it may be coded as an edge, but if the orientation content is broadly distributed then inhibitory mechanisms could provide a signal to sum orientations (and perhaps frequencies) to extract texture properties.

Work using natural images (Neri, 2014) and textures (Baker and Meese, 2014) has produced data that are broadly consistent with our results. Neri (2014) found evidence inhibition/suppression mechanisms when observers detected Gabors in noise that were either congruent/incongruent with the underlying orientation of natural scenes. He measured orientation tuning via the classification image technique and found orientation tuning and signatures of inhibitory mechanisms similar to those presented in our results (see **Figure 1G**). Baker and Meese (2014) used a contrast increment detection task and reverse correlation to measure the extent over which information is summed in visual space. Their reverse correlation results (see their **Figures 3G,H**) show the hallmarks of suppression beyond 5° of visual angle from fixation. Taken together the results above and our data provide converging lines of evidence for the use of inhibitory





mechanisms that adjust tuning in orientation and visual space.

## 5. CONCLUSION

The goal of this paper was to determine if the results we found in our previous work on spatial frequency summation (Taylor et al., 2009) extended to orientation summation using visual noise as a stimulus. We found that detection thresholds in human and ideal observers were proportional to the quarter-root of the number of spatial Fourier components in the stimulus. Hence, orientation summation, like spatial frequency summation, was nearly optimal across a wide range of bandwidths. However, unlike what we found with spatial frequency summation, our classification image results were inconsistent with a fixed channel model. Instead, our results suggest that the orientation bandwidth of the internal filter used to detect our stimuli was adjusted to match (albeit imperfectly) the orientation bandwidth of the stimulus. The classification images also show hallmarks of inhibition at uninformative spatial frequencies and orientations and lead to the hypotheses that human observers may detect noise stimuli by comparing the power in different portions of the power spectrum.

## REFERENCES

- Abbey, C., Eckstein, M., and Bochud, F. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. *Proc. SPIE* 3663, 284–295. doi: 10.1117/12.349653
- Abbey, C. K., and Eckstein, M. P. (2002). Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. *J. Vis.* 2, 66–78. doi: 10.1167/2.1.5
- Baker, D. H., and Meese, T. S. (2012). Zero-dimensional noise: the best mask you never saw. *J. Vis.* 12, 1–12. doi: 10.1167/12.10.20
- Baker, D. H., and Meese, T. S. (2014). Measuring the spatial extent of texture pooling using reverse correlation. *Vis. Res.* 97, 52–58. doi: 10.1016/j.visres.2014.02.004
- Brainard, D. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 443–446. doi: 10.1163/156856897X00357

- Burr, D., Morrone, C., and Maffei, L. (1981). Intra-cortical inhibition prevents simple cells from responding to textured visual patterns. *Exp. Brain Res.* 43, 455–458.
- Campbell, F., and Kulikowski, J. (1966). Orientational selectivity of the human visual system. *J. Physiol.* 187, 437.
- Campbell, F., Kulikowski, J., and Levinson, J. (1966). The effect of orientation on the visual resolution of gratings. *J. Physiol.* 187, 427.
- Carandini, M., and Ringach, D. (1997). Predictions of a recurrent model of orientation selectivity. *Vis. Res.* 37, 3061–3071. doi: 10.1016/S0042-6989(97)00100-4
- Dakin, S. C., Mareschal, I., and Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vis. Res.* 45, 3027–3049. doi: 10.1016/j.visres.2005.07.037
- De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res.* 22, 545–559. doi: 10.1016/0042-6989(82)90113-4
- Derrington, A., and Henning, G. (1989). Some observations on the masking effects of two-dimensional stimuli. *Vis. Res.* 29, 241–246. doi: 10.1016/0042-6989(89)90127-2
- DeValois, R., and DeValois, K. (1988). *Spatial Vision*. New York: Oxford University Press.
- Efron, B., and Tibshirani, R. (1994). *Introduction to the Bootstrap*. New York: Chapman & Hall.
- Gegenfurtner, K. R., and Kiper, D. C. (1992). Contrast detection in luminance and chromatic noise. *JOSA A* 9, 1880–1888. doi: 10.1364/JOSA.9.001880
- Girshick, A. R., Landy, M. S., and Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14, 926–932. doi: 10.1038/nn.2831
- Gold, J., Bennett, P., and Sekuler, A. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vis. Res.* 39, 3537–3560. doi: 10.1016/S0042-6989(99)00080-2
- Goris, R. L., Putzeys, T., Wagemans, J., and Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychol. Rev.* 120, 472. doi: 10.1037/a0033136
- Gosselin, F., and Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vis. Res.* 41, 2261–2271. doi: 10.1016/S0042-6989(01)00097-9
- Gowenlock, S., Taylor, C., Sekuler, A., and Bennett, P. (2009). The effect of aging on the orientational selectivity of the human visual system. *Vis. Res.* 49, 164–172. doi: 10.1016/j.visres.2008.10.004
- Graham, N. V. S. (1989). *Visual Pattern Analyzers*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195051544.001.0001
- Green, D. (1960a). “Auditory detection of a noise signal,” in *Signal Detection and Recognition by Human Observers*, ed J. Swets (New York: Wiley), 523–547.
- Green, D. (1960b). Auditory detection of a noise signal. *J. Acoust. Soc. Am.* 32, 121–131. doi: 10.1121/1.1907862
- Green, D., and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Heitger, F., Rosenthaler, L., Von Der Heydt, R., Peterhans, E., and Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vis. Res.* 32, 963–981. doi: 10.1016/0042-6989(92)90039-L
- Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vis. Res.* 27, 1029–1040. doi: 10.1016/0042-6989(87)90016-2
- Klein, S. A. (1992). “Channels: bandwidth, channel independence, detection vs. discrimination,” in *Channels in the Visual Nervous System: Neurophysiology, Psychophysics and Models*, ed B. Blum (London: Freund Publishing House), 11–27.
- Levi, D. M., Klein, S. A., and Chen, I. (2005). What is the signal in noise? *Vis. Res.* 45, 1835–1846. doi: 10.1016/j.visres.2005.01.020
- Levi, D. M., Klein, S. A., and Chen, I. (2008). What limits performance in the amblyopic visual system: seeing signals in noise with an amblyopic brain. *J. Vis.* 8, 1–23. doi: 10.1167/8.4.1
- Murray, R. (2011). Classification images: a review. *J. Vis.* 11, 1–25. doi: 10.1167/11.5.2
- Nachmias, J., Sansbury, R., Vassilev, A., and Weber, A. (1973). Adaptation to square-wave gratings: in search of the elusive third harmonic. *Vis. Res.* 13, 1335–1342. doi: 10.1016/0042-6989(73)90209-5
- Neri, P. (2014). Semantic control of feature extraction from natural scenes. *J. Neurosci.* 34, 2374–2388. doi: 10.1523/JNEUROSCI.1755-13.2014

- Pack, C. C., Livingstone, M. S., Duffy, K. R., and Born, R. T. (2003). End-stopping and the aperture problem: two-dimensional motion signals in macaque v1. *Neuron* 39, 671–680. doi: 10.1016/S0896-6273(03)00439-2
- Pelli, D. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366
- Pelli, D., and Farell, B. (1999). Why use noise? *J. Opt. Soc. Am. A* 16, 647–653. doi: 10.1364/JOSAA.16.000647
- Perkins, M., and Landy, M. (1991). Nonadditivity of masking by narrow-band noises. *Vis. Res.* 31, 1053–1065. doi: 10.1016/0042-6989(91)90209-N
- Quick, R., Hamerly, J., and Reichert, T. (1976). The absence of a measurable “critical band” at low suprathreshold contrasts. *Vis. Res.* 16, 351–355. doi: 10.1016/0042-6989(76)90195-4
- Ringach, D. (1998). Tuning of orientation detectors in human vision. *Vis. Res.* 38, 963–972. doi: 10.1016/S0042-6989(97)00322-2
- Ringach, D., Shapley, R., and Hawken, M. (2002). Orientation selectivity in macaque v1: diversity and laminar dependence. *J. Neurosci.* 22, 5639.
- Sceniak, M. P., Hawken, M. J., and Shapley, R. (2001). Visual spatial characterization of macaque v1 neurons. *J. Neurophysiol.* 85, 1873–1887.
- Schwartz, O., and Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4, 819–825. doi: 10.1038/90526
- Shirazi, M. (2004). Emergence of orientation-selective inhibition in the primary visual cortex: a bayes–markov computational model. *Biol. Cybern.* 91, 115–130. doi: 10.1007/s00422-004-0483-5
- Solomon, J. (2002). Noise reveals visual mechanisms of detection and discrimination. *J. Vis.* 2, 105–120. doi: 10.1167/2.1.7
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2003). Noise detection: bandwidth uncertainty and adjustable channels. *J. Vis.* 3, 9–9. doi: 10.1167/3.9.9
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2004). Noise detection: optimal summation of orientation information. *J. Vis.* 4, 50–50. doi: 10.1167/4.8.50
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2005). Noise detection: summation of high spatial frequency information. *J. Vis.* 5, 482–482. doi: 10.1167/5.8.482
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2006). Narrow-band channels optimally sum a broad band of spatial frequency information. *J. Vis.* 6, 115–115. doi: 10.1167/6.6.115
- Taylor, C. P., Bennett, P. J., and Sekuler, A. B. (2009). Spatial frequency summation in visual noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 26, B84–B93. doi: 10.1364/JOSAA.26.000B84
- Wandell, B. (1995). *Foundations of Vision*. New York: Wiley.
- Watson, A. B. (2000). Visual detection of spatial contrast patterns: evaluation of five simple models. *Opt. Express* 6, 12–33. doi: 10.1364/OE.6.000012
- Watson, A. B., and Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *JOSA A* 14, 2379–2391. doi: 10.1364/JOSAA.14.002379
- Wetherill, G., and Levitt, H. (1965). Sequential estimation of points on a psychometric function. *Br. J. Math. Stat. Psychol.* 18, 1–10. doi: 10.1111/j.2044-8317.1965.tb00689.x
- Wilson, H., McFarlane, D., and Phillips, G. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vis. Res.* 23, 873–882. doi: 10.1016/0042-6989(83)90055-X
- Yeshurun, Y., Carrasco, M., and Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: tests of the difference model. *Vis. Res.* 48, 1837–1851. doi: 10.1016/j.visres.2008.05.008

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 January 2014; paper pending published: 15 April 2014; accepted: 23 May 2014; published online: 12 June 2014.

Citation: Taylor CP, Bennett PJ and Sekuler AB (2014) Evidence for adjustable bandwidth orientation channels. *Front. Psychol.* 5:578. doi: 10.3389/fpsyg.2014.00578  
This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Taylor, Bennett and Sekuler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparison of informational vs. energetic masking effects on speechreading performance

Björn Lidestam<sup>1\*</sup>, Johan Holgersson<sup>1</sup> and Shahram Moradi<sup>2</sup>

<sup>1</sup> Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

<sup>2</sup> Linnaeus Centre HEAD, Swedish Institute for Disability Research, Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Mark W. Greenlee, University of Regensburg, Germany  
Eduardo Lugo, Visual Psychophysics and Perception Laboratory, Canada

## \*Correspondence:

Björn Lidestam, Department of Behavioral Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden  
e-mail: bjorn.lidestam@liu.se

The effects of two types of auditory distracters (steady-state noise vs. four-talker babble) on visual-only speechreading accuracy were tested against a baseline (silence) in 23 participants with above-average speechreading ability. Their task was to speechread high frequency Swedish words. They were asked to rate their own performance and effort, and report how distracting each type of auditory distracter was. Only four-talker babble impeded speechreading accuracy. This suggests competition for phonological processing, since the four-talker babble demands phonological processing, which is also required for the speechreading task. Better accuracy was associated with lower self-rated effort in silence; no other correlations were found.

**Keywords:** speech perception, cognition, speechreading, informational masking, energetic masking

## INTRODUCTION

In everyday speech perception, we hear speech clearly and without effort. Speech perception is usually pre-dominantly auditory. We might see the person talking, and this may help us perceive the speech more distinctly, especially if the speech signal is degraded or masked by noise (e.g., Hygge et al., 1992; Calvert et al., 1997; Moradi et al., 2013), but often the acoustic speech signal is enough for us to hear what is spoken. For most people, hearing speech is usually effortless and efficient. We are able to perceive a sufficient proportion of the speech sounds for comprehension of the speech signal.

Occasionally a speech signal will be masked by noise (i.e., sounds other than the voice of the person we are trying to hear). There are two main ways that noise can interfere with the speech signal. First, the noise can physically interfere with the speech signal (i.e., outside of the perceiver, in the acoustic environment). This is often referred to as *energetic masking* (Pollack, 1975). Second, the noise can perceptually interfere with the speech signal (i.e., inside the perceiver, in the perceptual process). This is often referred to as *informational masking* (Pollack, 1975; Watson et al., 1976).

Disentangling informational masking from energetic masking in auditory perception is difficult, as can be seen in the literature (e.g., Watson, 2005; Yost, 2006; Kidd et al., 2007). How can the detrimental effect of noise on speech perception be attributed to either informational masking, or to energetic masking (or to attentional allocation as a result of stimulus degradation)? Obviously, if an acoustic speech signal is presented together with an acoustic noise signal, there will necessarily be some degree of energetic masking. Elaborate study designs (such as that in Mattys et al., 2009) are required to dissociate the two types of masking.

The present study solved this problem by *not* presenting an acoustic speech signal, and by instead testing the effect of two closely matched types of noise (henceforth referred to as *auditory*

*distracters*) on visual-only speechreading. That is, there was no possibility of energetic masking, as there was no acoustic signal to mask, interfere with, or compete with. Any effects of the auditory distracter could therefore be attributed to either attentional or phonological processing, or to a combination of both.

In order to test whether there was a general effect on attention, a broadband *steady-state noise* (SSN) was used. As the SSN does not contain phonological information, its potential effect on speechreading performance (i.e., linguistic processing) is likely to be indirect. Specifically, more attentional resources might be needed for stream segregation, leaving less for the search of critical visual speech features when trying to make a lexical match, thereby lowering speechreading performance. Alternatively, SSN could improve speechreading performance via stochastic resonance, whereby a signal (such as auditory noise) in one modality can facilitate perception in another modality (see e.g., Harper, 1979; Manjarrez et al., 2007; Söderlund et al., 2007; Lugo et al., 2008; Ward et al., 2010; Tjan et al., 2013; Gleiss and Kayser, 2014).

In order to test whether there was an effect on phonological processing, segmented *four-talker babble* (4TB) was used. The 4TB was matched to the broadband SSN in terms of average sound intensity and frequency. The 4TB is speech, and so contains phonological information. Any effect of the 4TB on speechreading performance will therefore be more complex. There could be a general effect on attention and stream segregation, or facilitation from stochastic resonance, similar to effects of the SSN. There could also be an effect of competition for phonological processing capacity (i.e., identifying the babble sounds as speech and making lexical matches), while simultaneously having to decode the visual speech movements as phonemes in order to make a lexical match.

Lower speechreading performance from the 4TB condition compared to the SSN condition would indicate an effect on phonological processing. Both auditory distracter types contained equivalent levels of acoustic energy across the frequency

spectrum and should therefore affect general attention or induce stochastic resonance to a similar extent. However, the 4TB also requires phonological processing, whereas the SSN does not. Brungart and Simpson (2005) found that visual-only identification of a word was impeded only by simultaneous auditory presentation of another word (spoken by one talker). They therefore suggested that only *speech* as auditory distracters can impair visual-only speechreading (since other distracters yielded no effects). Furthermore, auditory distracters must according to Brungart and Simpson (2005) be presented simultaneously with the visual speech (as synchronicity reduced the impairment).

In the present study, we wanted our visual speech identification task to be as free from contextual cues as possible, as has been the case in most studies on auditory speech perception in noise. In studies on auditory speech perception, the standard case is to allow the participants to perceive relatively clearly what is being said; the acoustical speech signal is rich in information and can be effortlessly identified without contextual cues. This is usually *not* the case for visual-only speechreading, since the optical speech signal is poorly defined (as compared to standard acoustical speech signals). We wanted our speechreading task to be primarily processed bottom up. That is, we wanted it to be a context-free (or non-primed) visual speech identification task, and for two reasons. First, we wanted to have high external and ecological validity, that is, to make the speechreading task as similar to everyday speech perception as possible (e.g., like watching someone talk behind a window pane or seeing someone talk on TV with the sound turned off; in real life we usually do not get closed sets of response alternatives). Second, we wanted the speechreading task to demand high sensitivity to phonological features to allow phonemic–lexical matches, with little influence from top-down support, in order to maximize chances for the auditory distracters to disturb speech identification.

However, it is not possible to use such a bottom-up task with a normal population without obtaining floor effects, since optical speech signals are poorly defined. Most individuals do not perform above chance levels on visual speech decoding tasks unless there is strong contextual support for top-down inferences, such as from script (e.g., Samuelsson and Rönnerberg, 1993), topic (e.g., Hanin, 1988), emotional cues (e.g., Lidestam et al., 1999), or a closed set of response alternatives (e.g., coordinate response measure, Brungart and Simpson, 2005). Including strong contextual cues or having a closed set of response alternatives can improve speechreading performance to relatively high levels for a cross-section of normal-hearing participants. However, such improved accuracy is not necessarily the result of more efficient lexical processing. If sufficient contextual cues are available, it is possible that responses are based on post-lexical inferences rather than on actual lexical matches. Hence, a substantial proportion of the responses (made following the presentation of strong contextual cues) may reflect educated guesses (“off-line” responses) rather than improved perceptual accuracy (“on-line” responses). In order to maximize the chances for linguistic (phonemic–lexical matching) processing in visual speechreading, we screened a relatively large number of individuals, and used only the best-performing speechreaders in the actual experiment, asking them to speechread everyday words without contextual cues.

This study aimed to shed more light on informational masking (i.e., disturbed speech perception) by contrasting two different auditory distracters: speech (i.e., the 4TB) compared to SSN. The 4TB was a continuous stream of speech, and was therefore not presented synchronously with the target words, as was the case in Brungart and Simpson (2005). As a baseline condition, speechreading in silence (i.e., without auditory distracter) was used. Effects of SSN could only be attributed to general attentional processes, as SSN does not contain phonological information. On the other hand, 4TB, with speech as an auditory distracter, contains phonological information. Any difference between SSN and 4TB can therefore be attributed to impeded visual phonemic–lexical matching elicited by the 4TB distracter signal. A negative effect of either type of auditory distracter would suggest that synchronicity is not required to impair speechreading accuracy. A positive effect on speechreading performance would suggest facilitation from stochastic resonance.

A secondary purpose of the study was to examine how the auditory distracter conditions were subjectively experienced in terms of level of distraction, effect on performance, and effort, to validate the effects on speechreading accuracy.

Finally, this study aimed to test whether there were correlations between self-rated variables and speechreading performance, in order to aid interpretations of how attention and phonological processing were affected by the auditory distracter conditions.

## SCREENING TEST

### METHODS

#### Participants

A total of 147 students at Linköping University (90 women, 53 men, and 4 who did not divulge sex and age), aged 18–37 years ( $M = 21.6$  years,  $SD = 2.8$  years), volunteered to take part in the study.

#### Materials

The stimulus materials were video recordings of the best identified 30 words as used in the study by Lidestam and Beskow (2006). Half of the words were from a “visit to a doctor” script, and half were from a “visit to a restaurant” script. The recordings showed a man speaking one word at a time, with a neutral facial expression. The words consisted of three to seven letters (and phonemes), with one or two syllables. All words were rated as highly typical for their respective script. The presentation showed the talker’s face and shoulders, and no shadows obscured the mouth or speech movements. For a detailed description, see Lidestam and Beskow (2006).

#### Procedure

The screening test was conducted in lecture halls. The stimuli were presented with video projectors onto either one or two (if available) large screens. After written informed consent was obtained, the participants positioned themselves within the lecture hall in such a manner that the screen was easily visible. They were encouraged to sit so they would not be able to see other participants’ response sheets. After seating, they were provided with response sheets and pencils, and informed about the general purpose of the study. Specifically, the participants were informed that

the study was about speechreading and that this first part was a screening test for an experiment that would be more exhaustive and rewarded with a cinema ticket. It was made known that only the best speechreaders would be invited to take part in the main study, if they agreed to do so (participants indicated their willingness by checking a box on the response sheet).

The participants were instructed that their task was to speechread (without sound) the words spoken in two scripts: “a visit to a doctor” and “a visit to a restaurant.” It was stated clearly that there was no hidden agenda, and that it was important to try their best to guess and to respond to all stimuli. They were also informed that the responses did not need to be whole words, and that parts of words were preferred as responses over no response at all, but that if only a part of a word was rendered (e.g., a consonant), its position in the word should be indicated.

Stimuli were presented in two script blocks. Before presentation of each block, the respective scenario was presented with text on the screen. The words were then presented at a reasonable pace that allowed all participants to respond without undue stress. The screen was black in between presentation of the words. At the end of the screening sessions, the participants indicated whether they could be contacted for the experiment that would follow. In total, the screening session took 20 min. After the session, the participants were given the opportunity to ask questions and were offered refreshments.

The responses for phonetic correctness were scored on a whole word basis; that is, each word was scored dichotomously as either correct or incorrect. Omissions or inclusions of word endings with /t/ were disregarded (e.g., “normal” vs. “normalt” [normal vs. normally]; “dåligt” vs. “dålig” [bad vs. badly]).

## RESULTS AND DISCUSSION

Mean speechreading performance in the screening test was  $M = 2.2$  words ( $SD = 2.55$  words, range 0–12 words). Fifty-two percent of participants responded with zero or only one word correct. Out of the 147 participants in the screening test, 130 agreed to be contacted for the experiment. Their mean score was  $M = 2.2$  words ( $SD = 2.48$  words, range 0–12 words).

These results show that visual-only speechreading is a difficult task for most individuals. Just over half of the participants correctly identified 0 or 1 word out of 30. However, the top performers (the best 5%) could identify as many as one-third of the words (but this came as a surprise to them when told about their results). This shows that there is considerable variability in the population of normal-hearing young students with regard to speechreading ability.

## MAIN EXPERIMENT

### METHODS

#### Participants

All participants who achieved a total score of three or more on the screening test and who had indicated on the scoring sheets that they could be contacted for participation in the main experiment ( $n = 43$ ) were asked to participate. Potential participants were informed that normal hearing was a requirement, and that their participation would be rewarded with a cinema ticket. A total of 23 students (21 women and 2 men), with a mean age of

21.9 years ( $SD = 2.7$  years, range 19–31 years), participated in the experiment.

#### Materials

The stimuli were video recordings of a woman speaking a selection of the 5000 most common Swedish words in everyday use (according to the Swedish Parole corpus; Språkbanken, n.d.). The talker's face and shoulders were shown, and indirect lighting was used so that no shadows obscured the speech movements.

We wanted to use common, everyday Swedish words that were relatively easy to speechread, even without contextual cues. The words were therefore chosen according to the following criteria. First, each word had to be ranked among the 5000 highest frequency Swedish words according to the Parole corpus. Second, variation with regard to the number of syllables was considered; hence, words with one to five syllables were used. Third, the majority of the stimulus words contained consonants that are relatively easy to identify visually, and preferably in initial position. Before deciding which words to use, all candidate words were scored for visual distinctiveness according to whether any of the visually distinct consonants /f v b m p/ were part of the word, and a bonus score was given if the visually distinct consonant was in initial position (i.e., the first or second phoneme). The score was then normalized by dividing the sum of the scores for visually distinct consonants and bonus scores for initial position by the total number of phonemes in the word. A total of 180 words were chosen using this procedure. The words were divided into three different lists with 60 words in each. The lists were balanced in terms of: visual distinctiveness, word frequency (according to the Parole corpus), initial phoneme, and number of phonemes per word (Supplementary Material).

A Sony DCR-TRV950 video camera was used to record the stimuli to mini-DV tape in PAL standard at 25 frames per second. Each stimulus word was recorded twice and the best recording of each word was chosen. The recording was edited into separate QuickTime files, one per stimulus word, in H.264 video format at  $640 \times 480$  pixels. Only the video track was exported, in order to eliminate the risk of speech sound being presented. Each video file was edited so that the first frame was repeated for 25 frames (i.e., 1 s) before the actual playback of the video. (This was done in order to cue the participant to the presentation, and to minimize the risk of failure to play back at the correct frame rate due to processing demands, as video playback tends to lag within the first second when using standard software such as QuickTime for playback).

Each stimulus file was then edited into one new file per condition. The files for the baseline condition in silence were kept without sound, whereas each file for presentation in the SSN condition included a unique part of the SSN, and each file for the 4TB condition included a unique part of the 4TB.

The SSN was the stationary, speech-shaped broadband noise used in the Swedish Hearing in Noise Test (HINT; Hällgren et al., 2006), and has the same long-term average spectrum as the HINT sentences. The original file with the 4TB was 2 min in duration, and comprised recordings of two male and two female native Swedish talkers reading different paragraphs of a newspaper text. It was post-filtered to resemble the long-term average spectrum



of the HINT sentences (Ng et al., 2013). In order to prevent participants from directing their attention to the content of the 4TB sentence (which was a finding suggested by the pilot study), the file was cut up into approximately 0.5 s sections, and scrambled so that the order of sections 1, 2, 3, 4, 5, 6 became 1, 3, 2, 4, 6, 5, and so on. Pilot testing verified that this was well tolerated by participants. It also indicated that the stimuli no longer roused attention regarding content. There were no apparent clicks resulting from the editing. For a comparison of the long-term average spectrum of the two auditory distracter types, see **Figure 1**. For a comparison of the spectral-temporal contents of the two auditory distracter types over a segment of 1 s, see **Figure 2** (SSN) and **Figure 3** (4TB).

The apparatus for presentation included an Apple iMac 8.1 computer with a 2.4 GHz Intel Core Duo processor, 2 GB RAM, and an ATI Radeon HD 2400 XT with 128 MB VRAM. A 20-inch monitor (set at 800 × 600 pixels), Primax Soundstorm 57450 loudspeakers (capable of 80–18,000 Hz), and Tcl/Tk and QuickTimeTcl software were used to present the stimuli.

A Brüel and Kjær sound level meter type 2205 with a Brüel and Kjær 1 inch free-field microphone type 4117 were used to monitor

sound pressure levels of the auditory distracters. These were placed at the approximate position of the participants' ears. Both auditory distracter types had equivalent continuous A-weighted sound pressure levels (LAeq) of 61 dB (SSN range = 59.7–62 dB; 4TB range = 52.4–70 dB) for the 2 min measurement during which the entire auditory distracter files were presented.

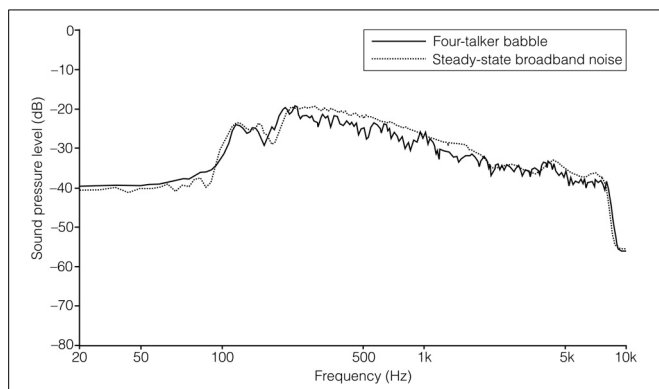
In order to examine how the auditory distracter conditions were subjectively experienced in terms of level of distraction, effect on performance, and effort, two questionnaires with 100 mm visual analogue scales were used. Scoring was calculated according to how many millimeters from the minimum (0 mm) the scale was ticked by the participants; hence maximum score was 100 mm.

### Procedure

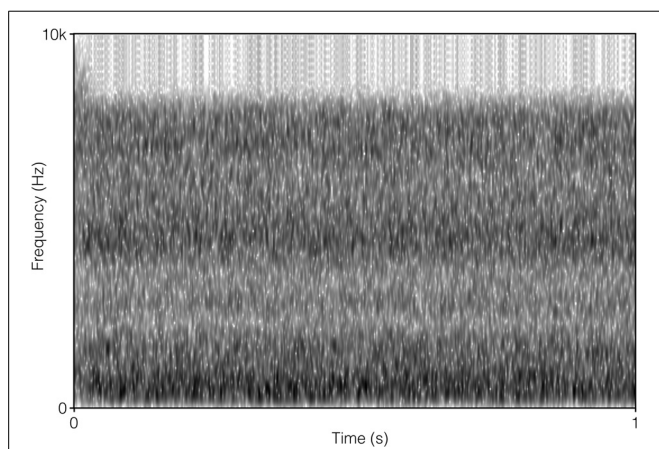
Each participant was seated in front of the monitor at a distance of approximately 60 cm. They were briefed about the general purpose of the study (i.e., they were informed that their task involved speechreading under three different sound conditions), and written informed consent was obtained. A response sheet with numbered lines for each presented stimulus was introduced, and the participant was instructed to respond to all presented words and encouraged to guess. Then a recording of a word that was not included in the actual experiment was presented, with the same auditory distracter condition as the participant started the experiment with, to familiarize the participant with the procedure.

The stimuli were presented one at a time; the speed of presentation was dictated by the pace of participant responding, but there was a maximum limit of 1 min (which never needed to be used). The screen turned white in the pause between stimuli. For all three conditions (i.e., silence, SSN, and 4TB), the sound continued during the pause (i.e., in the silent condition, the pause was silent too; in the SSN condition, the SSN continued during the pause; and in the 4TB condition, the 4TB continued during the pause).

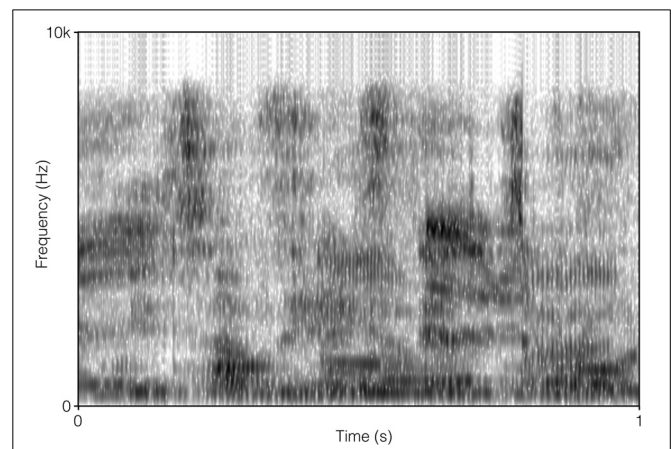
Scoring followed the procedure used in the screening test, such that the responses were dichotomously scored for phonetic correctness on a whole word basis.



**FIGURE 1 | The long-term average spectrum for the two auditory distracters.**



**FIGURE 2 | Spectrogram of 1 s of steady-state broadband noise.**



**FIGURE 3 | Spectrogram of 1 s of four-talker babble.**

After performing in each auditory distracter condition, the participant filled out the subjective experience questionnaire (which concerned experiences of each sound condition, self-ratings of performance, and which included some open-ended questions; see Supplementary Material). At the end of the experiment, the participants were awarded with a cinema ticket as reward for participation, and were given the opportunity find out more about the experiment. The experimental session took approximately 50 min to complete.

## Design

This study employed a within-groups design, with auditory distracter as the independent variable (three levels: silent, SSN, and 4TB), and speechreading accuracy as the dependent variable. A Latin-square design was used to determine the presentation orders of conditions (silence, SSN, and 4TB) and lists (1–3), so all experimental conditions and lists were combined and presented in all serial positions. Participants were randomized to presentation orders.

## RESULTS

### Effect of auditory distracter on speechreading accuracy

Auditory distracter significantly affected speechreading accuracy,  $F_{(2, 44)} = 11.19$ ,  $MSE = 6.21$ ,  $p < 0.001$ , partial  $\eta^2 = 0.34$ . Three *post-hoc* *t*-tests with Bonferroni corrected alpha ( $p < 0.17$ ) showed a significant difference between the 4TB and silence conditions,  $t_{(22)} = 2.98$ ,  $p = 0.007$ ,  $d = 0.62$ , and between the 4TB and SSN conditions,  $t_{(22)} = 4.65$ ,  $p < 0.001$ ,  $d = 0.97$ . There was no significant difference between the SSN and silence conditions,  $t_{(22)} = 1.71$ ,  $p = 0.101$ . In sum, only 4TB had an effect on speechreading accuracy, and this effect was negative (see Table 1).

Most error responses were words that included one or several correct phonemes, and one or several incorrect phonemes (about 75% of all responses belonged to this category). The second most common errors were words without any correct phoneme (and the majority of these errors were words with one or more phonemes which were easily visually confused with phonemes in the target word, such as /f/ instead of /v/ or /b/ instead of /p/). The third most common error was failure to respond with a proper word, such as only responding with a few letters as a part of a

word. Omissions (i.e., no response at all to a target word) constituted the least common cause of errors, with 8% of the total number of responses.

### Effects of auditory distracter on ratings of effort, distraction, and performance

Auditory distracter had a significant effect on participants' self-ratings of effort,  $F_{(2, 44)} = 3.40$ ,  $MSE = 3.30$ ,  $p = 0.042$ , partial  $\eta^2 = 0.13$ . Three *post-hoc* *t*-tests with Bonferroni correction ( $p < 0.17$ ) revealed only a tendency toward a significant difference between the silence and 4TB conditions,  $t_{(22)} = 2.32$ ,  $p = 0.03$ . The means (see Table 1) indicate that the speechreading task was generally perceived as effortful, and that speechreading in the 4TB condition was considered to be very effortful. Comparisons between the two auditory distracter conditions indicated no effect of auditory distracter type on rated distraction or self-rated performance. The mean ratings suggested that participants considered both types of auditory distracter to have impeded their performance to a considerable extent; both types of auditory distracter were rated as more toward "almost unbearable" than "not distracting at all."

### Correlations between speechreading accuracy, and ratings of effort, distraction, and performance

Table 2 presents the correlation results. The only significant correlation was between accuracy and self-rated effort in the silence condition,  $r_{(23)} = 0.44$ ,  $p < 0.05$ . Specifically, better performance was associated with lower effort ratings in the silent condition (high scores on the self-rating indicated low effort). However, there was no difference between the correlation coefficients for silence vs. 4TB.

## DISCUSSION

The present study showed that visual-only speechreading was only impeded by an auditory speech-based distracter, but not by noise itself. This implies that in order for the distracter to have an impact, it has to compete for phonological processing, which is required for identification of the visual speech signal. More "general" auditory distraction, such as the SSN stimuli used in this study, did not impede speechreading accuracy, in spite of that it was rated as very distracting by the participants. Competition for phonological processing (and following semantic processing) demands processing related to working memory, such that individuals with superior working-memory related capacities are less

**Table 1 | Descriptive statistics (means and standard deviations) for accuracy, and self-ratings of effort, performance, distraction, and effect of auditory distracter on performance.**

	Auditory distracter condition					
	Silence		SSN		4TB	
	M	SD	M	SD	M	SD
Accuracy (% correct)	14.8	6.6	16.9	6.2	11.2	5.9
Self-rated effort (0–100)	46.6	24.3	38.0	19.7	32.8	17.8
Self-rated performance (0–100)	42.5	15.1	44.1	18.3	37.9	16.1
Rated distraction of auditory distracter (0–100)	n.a.	n.a.	40.7	23.0	45.3	22.6
Rated effect of auditory distracter on performance (0–100)	n.a.	n.a.	40.2	16.7	32.7	18.0

**Table 2 | Pearson correlations between accuracy and ratings.**

	Auditory distracter condition		
	Silence	SSN	4TB
Self-rated effort	0.44*	0.00	0.17
Self-rated performance	0.11	0.22	−0.04
Rated distraction of auditory distracter	n.a.	−0.01	−0.15
Rated effect of auditory distracter on performance	n.a.	−0.01	0.09

\* $p < 0.05$ .

impeded by speech and speech-like distracters (e.g., Rönnberg et al., 2010; Zekveld et al., 2013). Markides (1989) showed an effect of classroom noise (including some speech sounds) on visual-only speechreading performance, but it is likely that the frequent and intermittent peaks of the noise (up to 97.5 dBA) interfered with attention as a result of their unpredictability and sheer sound pressure level—it is difficult not to be distracted by such loud sounds.

The participants in the present study were above-average speechreaders recruited among normally hearing students. Speechreading performance is positively correlated with aspects of working memory in this population (Lidestam et al., 1999). Therefore, the impediment effect on speechreading by a speech-based auditory distracter should be potentially even stronger on the majority of normally hearing individuals, since they generally have lower working-memory related capacities (Lidestam et al., 1999) and are more impeded by speech and speech-like distracters (e.g., Rönnberg et al., 2010; Zekveld et al., 2013). Individuals who are less proficient speechreaders also perceive the visual speech as very indistinct, making them even more disadvantaged (i.e., the weaker the percept, the easier to disrupt it).

The present study also showed that the auditory distracter signal does not need to be simultaneous in terms of onset relative to the visual speech signal, as suggested by Brungart and Simpson (2005). The auditory speech signal in the present study was four-speaker babble and was therefore more or less continuous.

Energetic masking can be ruled out as an explanation of impeded speech identification in this study, as there was no acoustic speech signal and hence no sound energy for the distracter signal to interfere with. Thus, the effect of the distracters on speechreading accuracy appears to have been purely “informational.”

No facilitation from either auditory distracter was found, but this should be further investigated in studies with more statistical power and higher sound pressure levels for SSN (in order for facilitation from stochastic resonance 70–80 dB is recommended; see e.g., Harper, 1979; Usher and Feingold, 2000; Manjarrez et al., 2007; Söderlund et al., 2007). The results from the present study suggest strongly that auditory speech distracters, such as 4TB, cannot facilitate speechreading, and it is unlikely that facilitation would occur under any sound pressure level. Many studies on auditory speech perception have found that speech and speech-like distracters, such as speech-shaped modulated noise, impede identification of speech targets (e.g., Festen and Plomp, 1990; Hygge et al., 1992; Hagerman, 2002; George et al., 2006; Zekveld et al., 2013).

As visual-only speech signals are generally poorly defined, almost any auditory distraction could potentially have a negative effect on the detection and identification of the subtle features of the speech movements involved. However, some previous studies failed to find effects, even of speech as distracter, on visual-only speechreading performance, except when the distracter signal was similar to the targets and presented synchronously (Brungart and Simpson, 2005; see also Lyxell and Rönnberg, 1993). In the Brungart and Simpson (2005) study, a coordinate response measure task was used in the condition where an effect of auditory distracters was found; this task has limited response alternatives.

Further, the distracter signal was a simultaneous auditory presentation of one talker speaking one of the few response alternatives to the visual target. Therefore, the task in that study can be assumed to have been more demanding in terms of attentional allocation and stream segregation (as there was only one talker, and the onset of the auditory distracter word was synchronized to the onset of the speech movements). For that reason, Brungart and Simpson's effect of phonological interference is more difficult to interpret than the findings of the present study. The generalizability to everyday speech perception of the results from the present study can also be claimed to be higher compared to the results in Brungart and Simpson's study, since everyday communication does not often provide such closed sets of response alternatives or situations resembling coordinate response measure tasks.

The hypothesis that average and below-average speechreaders should be more disturbed by auditory speech distracters, compared to above-average speechreaders, would require a highly structured task, such as a coordinate response measure task or stimuli that are extremely visually well defined, however. Floor effects would be difficult to avoid otherwise: if performance is at the floor at baseline it cannot decrease.

The only significant correlation found was between speechreading accuracy and self-rated effort in the silent condition (i.e., without auditory distracter). This finding may indicate that segregating the speech (the speech movements, the phonological information that the speech movements elicit, or both) from the distracter signal (i.e., the SSN or 4TB) increased the cognitive load, which made the ratings less accurate. That is, it is possible that there was not enough cognitive spare capacity to accurately rate own effort after segregating speech from noise, which would mean that the task was more cognitively demanding than realized by the participants. This explanation is in line with the conclusions from studies suggesting that segregating input from different signal sources requires cognitive effort (e.g., Mishra et al., 2013; Zekveld et al., 2013).

## ACKNOWLEDGMENTS

This research was funded by the Swedish Research Council (grant number 2006–6917). Two anonymous reviewers provided valuable comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00639/abstract>

## REFERENCES

- Brungart, D., and Simpson, B. (2005). Interference from audio distracters during speechreading. *J. Acoust. Soc. Am.* 118, 3889–3902. doi: 10.1121/1.2126932
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., et al. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593
- Festen, J. M., and Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J. Acoust. Soc. Am.* 88, 1725–1736. doi: 10.1121/1.400247
- George, E. L., Festen, J. M., and Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing impaired listeners. *J. Acoust. Soc. Am.* 120, 2295–2311. doi: 10.1121/1.2266530

- Gleiss, S., and Kayser, C. (2014). Acoustic noise improves visual perception and modulates occipital oscillatory states. *J. Cogn. Neurosci.* 26, 699–711. doi: 10.1162/jocn\_a\_00524
- Hagerman, B. (2002). Speech recognition in slightly and fully modulated noise for hearing impaired subjects. *Int. J. Audiol.* 41, 321–329. doi: 10.1019/14992020209090406
- Hällgren, M., Larsby, B., and Arlinger, S. (2006). A Swedish version of the Hearing In Noise Test (HINT) for measurement of speech recognition. *Int. J. Audiol.* 45, 227–237. doi: 10.1080/14992020500429583
- Hanin, L. (1988). *The Effects of Experience and Linguistic Context on Speechreading*. Doctoral dissertation, The City University Graduate School, New York, NY.
- Harper, D. W. (1979). Signal detection analysis of effect of white noise intensity on sensitivity to visual flicker. *Percept. Mot. Skills* 48, 791–798. doi: 10.2466/pms.1979.48.3.791
- Hygge, S., Rönnberg, J., Larsby, B., and Arlinger, S. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *J. Speech Hear. Res.* 35, 208–215.
- Kidd, G. Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2007). "Informational masking," in *Springer Handbook of Auditory Research*, Vol. 29: *Auditory Perception of Sound Sources*, eds W. A. Yost, A. N. Popper, and R. R. Fay (New York, NY: Springer), 143–189.
- Lidestam, B., and Beskow, J. (2006). Visual phonemic ambiguity and speechreading. *J. Speech. Lang. Hear.* 49, 835–847. doi: 10.1044/1092-4388(2006/059)
- Lidestam, B., Lyxell, B., and Andersson, G. (1999). Speech-reading: Cognitive predictors and displayed emotion. *Scand. Audiol.* 28, 211–217. doi: 10.1080/010503999424644
- Lugo, E., Doti, R., and Faubert, J. (2008). Ubiquitous crossmodal stochastic resonance in humans: auditory noise facilitates tactile, visual and proprioceptive sensations. *PLoS ONE* 3:e2860. doi: 10.1371/journal.pone.0002860
- Lyxell, B., and Rönnberg, J. (1993). The effects of background noise and working memory capacity on speechreading performance. *Scand. Audiol.* 22, 67–70. doi: 10.1019/01050399309046021
- Manjarrez, E., Mendez, I., Martinez, L., Flores, A., and Mirasso, C. R. (2007). Effects of auditory noise on the psychophysical detection of visual signals: Cross-modal stochastic resonance. *Neurosci. Lett.* 415, 231–236. doi: 10.1016/j.neulet.2007.01.030
- Markides, A. (1989). Background noise and lip-reading ability. *Br. J. Audiol.* 23, 251–253. doi: 10.1019/03005368909076507
- Mattys, S., Brooks, J., and Cooke, M. (2009). Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cogn. Psychol.* 59, 203–243. doi: 10.1016/j.cogpsych.2009.04.001
- Mishra, S., Lunner, T., Stenfelt, S., and Rudner, M. (2013). Seeing the talker's face supports executive processing of speech in steady state noise. *Front. Syst. Neurosci.* 7:96. doi: 10.3389/fnsys.2013.00096
- Moradi, S., Lidestam, B., and Rönnberg, J. (2013). Gated audiovisual speech identification in silence versus noise: effects on time and accuracy. *Front. Psychol.* 4:359. doi: 10.3389/fpsyg.2013.00359
- Ng, E. H. N., Rudner, M., Lunner, T., Syskind Pedersen, M., and Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int. J. Audiol.* 52, 433–441. doi: 10.1019/14992027.2013.776181
- Pollack, I. (1975). Auditory informational masking. *J. Acoust. Soc. Am.* 57, S5. doi: 10.1121/1.1995329
- Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. A. (2010). When cognition kicks in: working memory and speech understanding in noise. *Noise Health* 12, 263–269. doi: 10.4103/1463-1741.70505
- Samuelsson, S., and Rönnberg, J. (1993). Implicit and explicit use of scripted constraints in lip-reading. *Eur. J. Cogn. Psychol.* 5, 201–233. doi: 10.1080/09541449308520116
- Söderlund, S., Sikström, S., and Smart, A. (2007). Listen to the noise: noise is beneficial for cognitive performance in ADHD. *J. Child Psychol. Psychiatry* 48, 840–847. doi: 10.1111/j.1469-7610.2007.01749.x
- Språkbanken (n.d.). *PAROLE Corpus*. Available online at: <http://spraakbanken.gu.se/parole/>
- Tjan, B. S., Chao, E., and Bernstein, L. E. (2013). A visual or tactile signal makes auditory speech detection more efficient by reducing uncertainty. *Eur. J. Neurosci.* 39, 1323–1331. doi: 10.1111/ejn.12471
- Usher, M., and Feingold, M. (2000). Stochastic resonance in the speed of memory retrieval. *Biol. Cybern.* 83, L11–L16. doi: 10.1007/PL00007974
- Ward, L. M., MacLean, S. E., and Kirschner, A. (2010). Stochastic resonance modulates neural synchronization within and between cortical sources. *PLoS ONE* 5:e14371. doi: 10.1371/journal.pone.0014371
- Watson, C. S. (2005). Some comments on informational masking. *Acta Acoust.* 91, 502–512.
- Watson, C. S., Kelly, W. J., and Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *J. Acoust. Soc. Am.* 60, 1176–1186. doi: 10.1121/1.381220
- Yost, B. (2006). "Informational masking: what is it?," in *Paper Presented at the 2006 Computational and Systems Neuroscience (Cosyne) Meeting*. Available online at: [www.isr.umd.edu/Labs/NSL/Cosyne/Yost.htm](http://www.isr.umd.edu/Labs/NSL/Cosyne/Yost.htm)
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., and Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 134, 2225–2234. doi: 10.1121/1.4817926

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 March 2014; accepted: 04 June 2014; published online: 24 June 2014.

Citation: Lidestam B, Holgersson J and Moradi S (2014) Comparison of informational vs. energetic masking effects on speechreading performance. *Front. Psychol.* 5:639. doi: 10.3389/fpsyg.2014.00639

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Lidestam, Holgersson and Moradi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Effect of luminance noise on the object frequencies mediating letter identification

Cierra Hall<sup>1,2</sup>, Shu Wang<sup>1,3</sup>, Reema Bhagat<sup>1</sup> and J. Jason McAnany<sup>1,2,4</sup> \*

<sup>1</sup> Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL, USA

<sup>2</sup> Department of Bioengineering, University of Illinois at Chicago, Chicago, IL, USA

<sup>3</sup> Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA

<sup>4</sup> Department of Psychology, University of Illinois at Chicago, Chicago, IL, USA

## Edited by:

Jocelyn Faubert, Université de Montréal, Canada

## Reviewed by:

Caroline Blais, Université du Québec en Outaouais, Canada

Frederic J. A. M. Poirier, Université de Montréal, Canada

## \*Correspondence:

J. Jason McAnany, Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, 1855 West Taylor Street, Chicago, IL 60612, USA  
e-mail: jmcana1@uic.edu

**Purpose:** To determine if the same object frequency information mediates letter contrast threshold in the presence and absence of additive luminance noise (i.e., “noise-invariant processing”) for letters of different size.

**Methods:** Contrast thresholds for Sloan letters ranging in size from 0.9 to 1.8 log MAR were obtained from three visually normal observers under three paradigms: (1) high- and low-pass Gaussian filtered letters were presented against a uniform adapting field; (2) high- and low-pass Gaussian filtered letters were presented in additive white luminance noise; and (3) unfiltered letters were presented in high- and low-pass Gaussian filtered luminance noise. A range of high- and low-pass filter cutoffs were used to limit selectively the object frequency content of the letters (paradigms 1 and 2) or noise (paradigm 3). The object frequencies mediating letter identification under each paradigm were derived from plots of log contrast threshold vs. log filter cutoff frequency.

**Results:** The object frequency band mediating letter identification systematically shifted to higher frequencies with increasing log MAR letter size under all three paradigms. However, the relationship between object frequency and letter size depended on the paradigm under which the measurements were obtained. The largest difference in object frequency among the paradigms was observed at 1.8 log MAR, where the addition of white noise nearly doubled the center frequency of the band of object frequencies mediating letter identification, compared to measurements made in the absence of noise.

**Conclusion:** Noise can affect the object frequency band mediating letter contrast threshold, particularly for large letters, an effect that is likely due to strong masking of the low frequency letter components by low frequency noise checks. This finding indicates that noise-invariant processing cannot necessarily be assumed for large letters presented in white noise.

**Keywords:** visual noise, letter identification, contrast sensitivity, optotype, object spatial frequency, retinal spatial frequency

## INTRODUCTION

Letter optotypes are commonly used as test targets in basic studies of visual performance as well as in the clinical evaluation of visual function. An important consideration in the use of letter targets is that their Fourier spectra contain a broad range of object spatial frequencies, designated in cycles per letter (cpl; Parish and Sperling, 1991; Polder, 2003). Although visual sensitivity for spatially broad-band letter optotypes could potentially be based on any of the object frequencies contained in the letter, studies have shown that only a narrow band of object frequencies mediates contrast sensitivity (Alexander et al., 1994; Solomon and Pelli, 1994; Chung et al., 2002; Majaj et al., 2002; McAnany and Alexander, 2008; Oruc and Landy, 2009) and visual acuity (Anderson and Thibos, 1999). Furthermore, the narrow band of object frequencies that mediates performance depends on letter size, such that higher object frequencies (i.e., the edges of the letter) are used for larger letter

sizes, whereas lower object frequencies are used for smaller letter sizes (Chung et al., 2002; Majaj et al., 2002; McAnany and Alexander, 2008; Oruc and Landy, 2009; Alexander and McAnany, 2010; McAnany et al., 2011).

The standard approach for studying the object frequency information mediating visual acuity and contrast sensitivity for letters has been to remove or mask selected object frequencies contained in the letter, and then measure the effect on performance. That is, if removing a range of object frequencies does not affect performance, then those frequencies must not be necessary for the task. Conversely, if removing a range of object frequencies impairs performance, then those frequencies must be useful for performing the task. Two distinct approaches based on this logic have been used to identify the object frequencies mediating letter contrast sensitivity: a letter filtering approach (Alexander et al., 1994; Chung et al., 2002; McAnany and Alexander, 2008;



Alexander and McAnany, 2010) and a noise masking approach (i.e., “critical-band noise masking”; Solomon and Pelli, 1994; Majaj et al., 2002; Oruc and Landy, 2009). The former approach involves selectively removing object frequencies from the letter by spatial filtering, whereas critical-band noise masking attenuates the usefulness of selected object frequencies by masking them with spatially filtered luminance noise. Despite differences in approach, previous studies that have examined the effect of letter size on the object frequencies mediating contrast sensitivity are in good agreement. For example, the data of Chung et al. (2002), who used band-pass filtered letters, indicate that a linear function with a slope of approximately 1/3 describes the relationship between log object frequency and log letter size, for letters sizes of approximately 0.1–1.4 log MAR. Similarly, the data of Majaj et al. (2002), who used critical-band noise masking, indicate that a linear relationship with a slope of approximately 1/3 can describe the relationship between log object frequency and log letter stroke frequency, for letter sizes of approximately 0.3–2.8 log MAR.

Studies that employ visual noise as a tool to assess the object frequency information mediating contrast sensitivity or as a tool to assess visual function in patient populations (Nordmann et al., 1992; Yates et al., 1998; Levi and Klein, 2003; Pelli et al., 2004; Xu et al., 2006; Huang et al., 2007; McAnany et al., 2013) typically assume noise-invariant processing: i.e., the same mechanism and processing strategy are used in the absence and presence of noise. However, the addition of white luminance noise can alter the visual pathway mediating sensitivity under certain conditions, biasing processing from the magnocellular visual pathway towards the parvocellular visual pathway (McAnany and Alexander, 2009, 2010). Additionally, previous work has shown that higher object frequencies mediate contrast sensitivity under conditions biased toward the parvocellular pathway (McAnany and Alexander, 2008). Consequently, the addition of luminance noise might affect the object frequency band mediating letter identification. If noise affects the object frequencies mediating letter identification, then the interpretation of clinical tests that assess performance in noise (Pelli and Hoepner, 1989; Pelli et al., 2004) and basic studies of the information mediating letter identification could be complicated.

Given that the validity of noise-invariant processing in letter contrast sensitivity tasks has not been tested, the present study determined the effects of luminance noise on the object frequencies mediating contrast sensitivity for letters across a range of sizes. Estimates of object frequency were determined by low- and high-pass spatial filtering letters from the Sloan set. Letters were either presented against a uniform adapting field or in the presence of white additive luminance noise. The effects of luminance noise on object frequency were also assessed by measuring the object frequencies mediating letter identification using high- or low-pass filtered noise with the critical-band noise masking paradigm.

## MATERIALS AND METHODS

### OBSERVERS

Three of the authors (ages 22, 24, and 34 years) served as subjects. All had normal best-corrected visual acuity assessed with the ETDRS distance visual acuity chart and normal and contrast

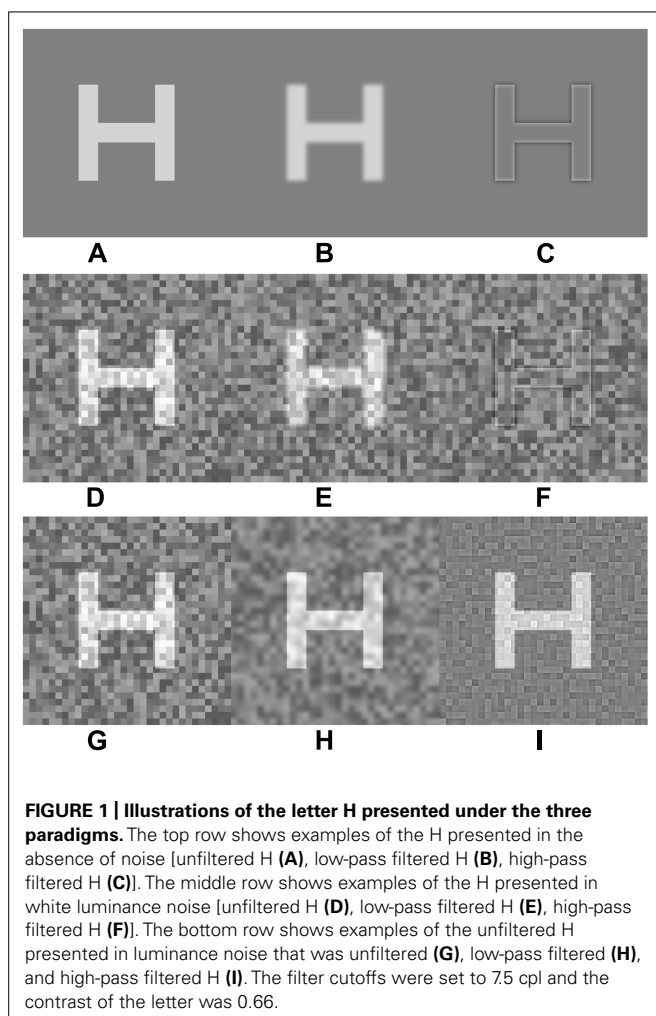
sensitivity assessed with the Pelli-Robson contrast sensitivity chart. The experiments were approved by an institutional review board at the University of Illinois at Chicago and the study adhered to the tenets of the Declaration of Helsinki.

### APPARATUS AND STIMULI

All stimuli were generated using a PC-controlled Cambridge Research Systems ViSaGe stimulus generator and were displayed on a Mitsubishi Diamond Pro (2070) CRT monitor with a screen resolution of  $1024 \times 768$  and a 100-Hz refresh rate. The monitor, which was the only source of illumination in the room, was viewed monocularly through a phoropter with the observer's best refractive correction. The luminance values used to generate the stimuli were determined by the ViSaGe linearized look-up table, which were verified by measurements made with a Minolta LS-110 photometer.

The test stimuli consisted of a set of ten Sloan letters (C, D, H, K, N, O, R, S, V, Z) that was constructed according to published guidelines (NAS-NRC, 1980). The Sloan letters were either unfiltered or spatially high- or low-pass filtered with a set of two-dimensional Gaussian filters. The object frequency cutoffs of the filters ranged from 0.9 to 21.0 cpl in 10 steps spaced approximately 0.15 log units apart. **Figure 1** presents examples of an unfiltered letter (**Figure 1A**), a low-pass-filtered letter (**Figure 1B**), and a high-pass-filtered letter (**Figure 1C**). The letters were of positive contrast (letter luminance higher than the adapting field luminance) and were presented at four different sizes, equivalent to 0.9, 1.2, 1.5, and 1.8 log MAR (minimum angle of resolution, where smaller values of log MAR correspond to smaller letters). This range was used in previous studies (McAnany and Alexander, 2008; Alexander and McAnany, 2010) and includes the letter size used for the Pelli-Robson contrast sensitivity chart for the standard 1 m test distance (1.5 log MAR). The letters were presented for an unlimited duration at the center of a  $50 \text{ cd/m}^2$  adapting field that subtended  $10.7^\circ$  horizontally and  $8.0^\circ$  vertically.

Letters were presented either in the absence of noise (**Figures 1A–C**) or in the presence of luminance noise (**Figures 1D–I**). The same letter targets shown in the absence of noise (first row) are shown in the presence of white additive luminance noise in the second row (**Figures 1D–F**). The bottom row of **Figure 1** provides examples of the stimuli used in the critical-band noise masking experiments. Examples of an unfiltered letter are shown in white noise (**Figure 1G**), in low-pass filtered noise (**Figure 1H**), and in high-pass filtered noise (**Figure 1I**). The noise field covered an area that was approximately 1.5 times larger than the letter and consisted of independently generated square checks with luminances drawn randomly from a uniform distribution with a root-mean-square (rms) contrast of 0.18. The mean luminance of the noise field was equal to that of the adapting field ( $50 \text{ cd/m}^2$ ). The size of the noise checks was scaled with letter size such that there were always 15 noise checks per letter (six checks per letter cycle, as each letter contains 2.5 cycles), which maintains a constant signal-to-noise ratio (SNR) across different letter sizes. Previous work has shown that a minimum of four checks per cycle are needed to ensure that the noise is effectively white at all letter sizes (Kukkonen et al., 1995). The value of six checks per letter cycle used in the present study is consistent with that used by others (Pelli



et al., 2004). The noise spectral density ranged from  $6 \times 10^{-6} \text{ deg}^2$  at the smallest check size to  $4 \times 10^{-4} \text{ deg}^2$  for the largest check size. The static noise field was presented synchronously with the target for an unlimited duration, such that the onset and offset of the target and noise was simultaneous.

The contrast (C) of the letters was defined as Weber contrast:

$$C = (L_L - L_B)/L_B \quad (1)$$

where  $L_L$  is the luminance of the letter and  $L_B$  is the background luminance. Because the contrast of complex images is difficult to define (Peli, 1990), a relative definition of contrast was used to characterize the filtered letters, as in previous studies (Chung et al., 2002; McAnany and Alexander, 2008, 2010). That is, when the contrast of the original unfiltered letter was 1.0, the filtered image was assigned a relative contrast of 1.0 without rescaling.

## PROCEDURE

A brief warning tone signaled the start of each stimulus presentation. On each trial, a single letter was selected at random from the Sloan set and presented. The observer's task was to identify the letter verbally, which was entered by the experimenter. No feedback was given. All three observers were familiar with the

Sloan set and only letters from the Sloan set were accepted as valid responses. Contrast threshold for letter identification was obtained using a 10-alternative forced-choice staircase procedure. An initial estimate of threshold was obtained by presenting a letter at a suprathreshold contrast level and then decreasing the contrast by 0.3 log units until an incorrect response was recorded. After this initial search, log contrast threshold was determined using a two-down, one-up decision rule, which provides an estimate of the 76% correct point on a psychometric function (Garcia-Perez, 1998). Each staircase continued until 16 reversals had occurred, and the mean of the last 6 reversals was taken as contrast threshold. Excluding the initial search, the staircase length was typically 35–40 trials, which produced stable measurements. In one testing session, a letter size and a paradigm (filtered letter in the absence of noise, filtered letter in the presence of noise, unfiltered letter in filtered noise) were selected pseudorandomly for testing. All cutoff object frequencies for both high-pass and low-pass filtered letters (or filtered noise) were tested in a pseudorandom order within a session.

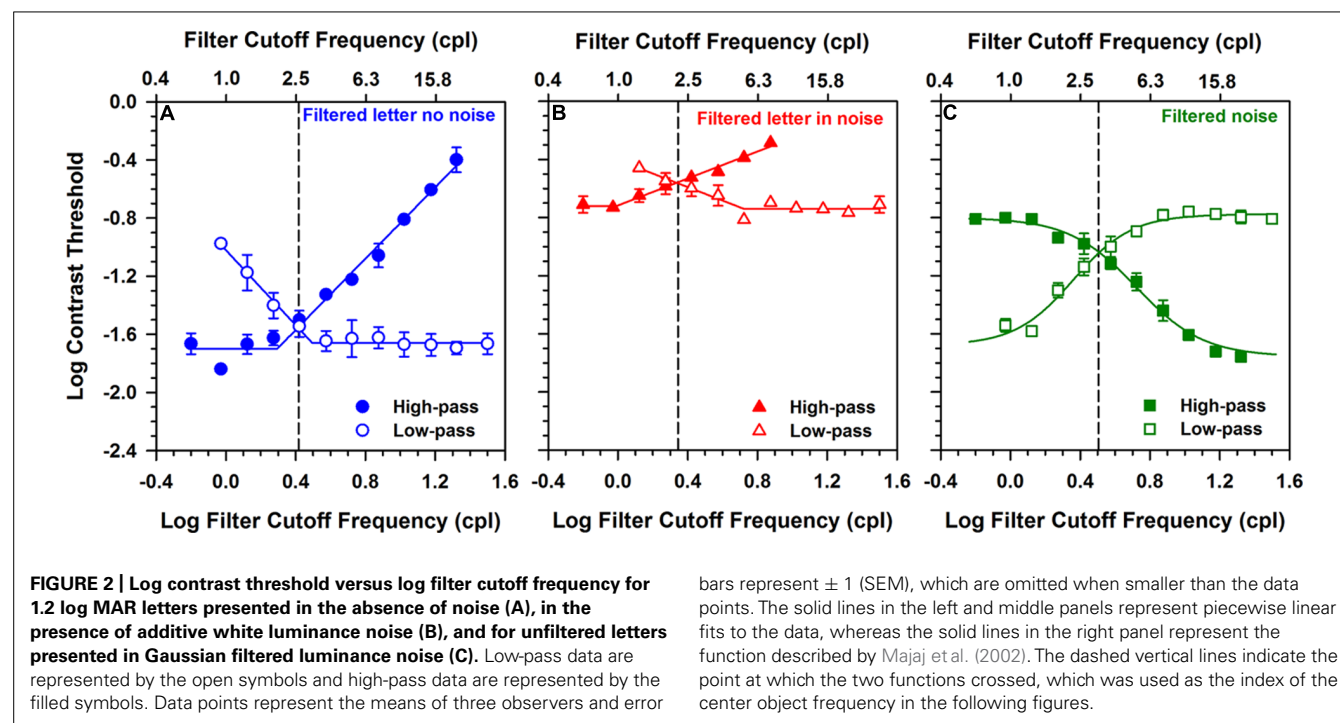
## RESULTS

### CONTRAST THRESHOLD FOR LOW- AND HIGH-PASS FILTERED LETTERS

Figures 2A,B show log contrast threshold for letters that were either high-pass Gaussian filtered (filled symbols) or low-pass Gaussian filtered (open symbols). These measurements were made for a letter size equivalent to 1.2 log MAR that was either presented in the absence of noise (Figure 2A) or in white luminance noise (Figure 2B). Each data point represents the mean contrast threshold value for the three subjects and the error bars are  $\pm 1$  standard error of the mean (SEM). In Figures 2A,B, the leftmost data points (filled circle and filled triangle, respectively) and rightmost data points (open circle and open triangle, respectively) represent contrast threshold for letters that were minimally filtered. The other data points represent the effect of successively changing the cutoff frequency of the filter to remove either the low object frequencies or the high object frequencies.

For the filtered letter data in Figures 2A,B, there was a region over which threshold was independent of filter cutoff and a second region over which log contrast threshold increased or decreased linearly with log filter cutoff. In order to derive the object frequency range that is used for letter identification, the data were fit piecewise with two linear functions using a least-squares criterion: one region was constrained to have a slope of 0, and the slope of the second region was unconstrained. The high-pass and low-pass functions in each plot were fit separately and are represented by the solid lines in Figures 2A,B. The cutoff object frequency at which the functions crossed (indicated by the vertical dashed lines) was taken as an index of the center of the object frequency region mediating letter identification. This point, which was also used in previous reports (McAnany and Alexander, 2008; Alexander and McAnany, 2010), represents approximately equal elevations of log contrast threshold, compared to the threshold values obtained with minimally filtered letters.

Log contrast thresholds for filtered letters measured in the absence of noise (Figure 2A) and in the presence of white noise (Figure 2B) differed substantially. That is, the functions measured in the presence of white noise were shifted vertically by approximately 1 log unit. This finding is expected, as high external



noise levels are known to elevate contrast threshold substantially. The center object frequency was similar under both conditions (approximately 2.6 cpl in the absence of noise and 2.2 cpl in the presence of noise). Thus, the functions obtained in the absence and presence of white noise were primarily shifted vertically, with minimal horizontal shift.

The range of useful frequencies mediating contrast threshold (i.e., bandwidth) can also be derived from the plots shown in **Figures 2A,B**. The bandwidth was calculated as the full width at half-height, where half-height was obtained by averaging the minimum and crossing point thresholds. The mean bandwidth was 0.67 octaves for the letters presented in the absence of noise (**Figure 2A**) and 1.1 octaves for letters presented in white noise (**Figure 2B**). The width of the band of object frequencies has been reported to be between 1 and 3 octaves (Chung et al., 2002; Majaj et al., 2002), with minimal dependence on letter size. Thus, the bandwidth measured in white noise is within the previously reported range, whereas the bandwidth measured in the absence of noise is somewhat narrower than previous reports.

#### CONTRAST THRESHOLD FOR LETTERS PRESENTED IN LOW- AND HIGH-PASS FILTERED NOISE

**Figure 2C** shows log contrast threshold for unfiltered letters presented in either high-pass Gaussian filtered (filled symbols) or low-pass Gaussian filtered (open symbols) white noise. The letter size was equivalent to 1.2 log MAR and each data point represents the mean contrast threshold value for the three subjects, with error bars representing  $\pm 1$  SEM. The leftmost point for the high-pass function (filled square) and rightmost data point for the low-pass function (open square) represent thresholds measured in noise that was minimally filtered (i.e., nearly white). The other data

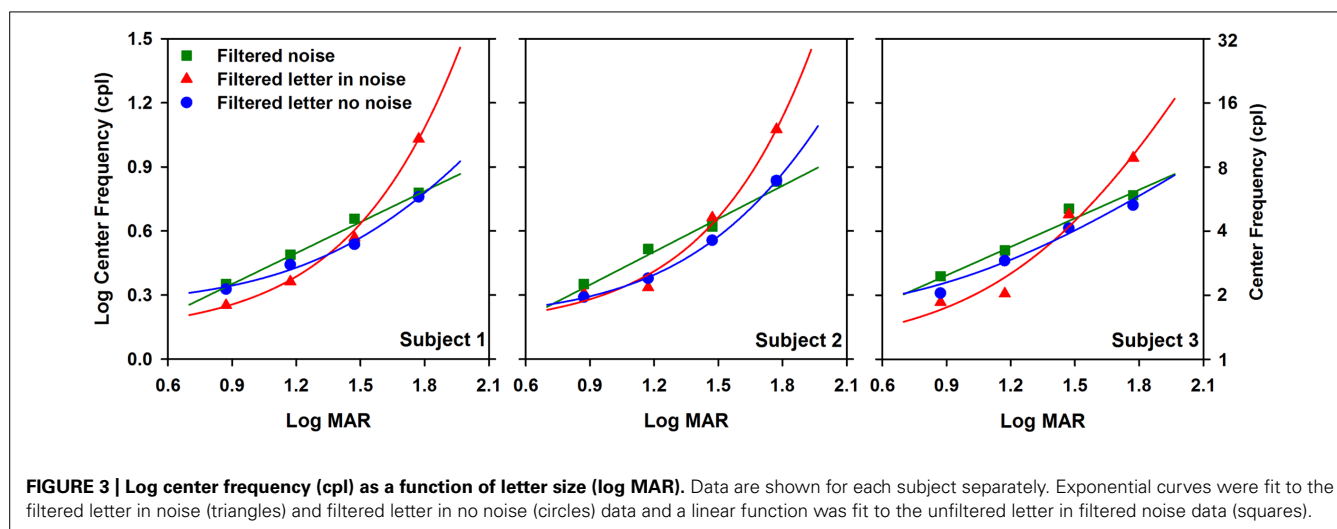
points represent the effect of successively changing the filter cutoff to remove object frequencies from the noise.

The data in **Figure 2C** were well fit by sigmoidal functions described previously and fit to similar data (Majaj et al., 2002; Oruc and Landy, 2009). The high-pass and low-pass functions were fit separately and are represented by the solid lines in **Figure 2C**. The crossing point of the fitted functions was taken as an index of the center of the range of frequencies mediating letter identification, to maintain consistency with the approach used in **Figures 2A,B**. Based on this definition, the center object frequency for the letters measured in filtered noise (3.2 cpl) was somewhat higher than the center frequencies in the absence of noise (2.6 cpl) or for letters in white noise (2.2 cpl). The center object frequency was also determined by calculating the derivatives of the sigmoidal curves, an approach described elsewhere (Majaj et al., 2002; Oruc and Landy, 2009). Estimates of the center object frequency based on the mean of the low- and high-pass derivatives was 3.5 cpl.

The bandwidth for letters presented in filtered noise (**Figure 2C**) was calculated using the same procedure described for the bandwidth calculations for **Figures 2A,B**. The full width at half-height of the data shown in **Figure 2C** was 2.1 octaves. This value is larger than that for filtered letters presented in the presence and absence of white noise.

#### CENTER OBJECT FREQUENCY AS A FUNCTION OF LETTER SIZE

The analysis illustrated in **Figure 2** was performed on the data obtained at each of the other three letter sizes, with the results shown in **Figure 3**. **Figure 3** shows log center object frequency for the three subjects as a function of log MAR letter size. The center object frequencies were based on the crossing points of the fits to the data, as described in **Figure 2**. Measurements are shown for



**FIGURE 3 | Log center frequency (cpl) as a function of letter size (log MAR).** Data are shown for each subject separately. Exponential curves were fit to the filtered letter in noise (triangles) and filtered letter in no noise (circles) data and a linear function was fit to the unfiltered letter in filtered noise data (squares).

filtered letters in the absence of noise (circles), filtered letters in the presence of white noise (triangles), and for unfiltered letters in filtered noise (squares). Data for filtered letters in the absence and presence of white noise were fit with exponential functions that transitioned from a slope of 0 for small letters to a positive slope for large letters. Data for the unfiltered letters presented in filtered noise were fit with a linear regression line, in accordance with previous reports (Majaj et al., 2002; Oruc and Landy, 2009). As can be seen from comparing the three panels, the results were highly consistent for the three subjects.

The relationship between object frequency and letter size was not identical for letters in the absence and presence of white noise. Specifically, center object frequency increased as size increased for both paradigms, but the exponential increase in center frequency for letters in white noise was greater than that for letters in the absence of noise. The largest difference between the functions measured in the presence and absence of white noise was at the largest size, where the object frequencies mediating letter identification in the presence of white noise were a factor of 1.75 higher, on average, than those measured in the absence of noise. For both paradigms, the slope of the function began to approach zero for small letter sizes, indicating that a similar constant band of frequencies mediated contrast threshold in the absence and presence of white noise for small letters. In comparison, log object frequency increased linearly with log letter size for the measurements made in filtered noise. The object frequencies mediating letter identification in filtered noise tended to be slightly, but systematically, higher than those measured in the absence of noise.

A two-way repeated measures analysis of variance (ANOVA) was performed to compare the object frequencies measured under the three paradigms at the four letter sizes. The ANOVA indicated significant main effects of paradigm [ $F(2,12) = 9.7, p < 0.05$ ] and letter size [ $F(3,12) = 125.4, p < 0.05$ ]. Additionally, there was a significant interaction between paradigm and size [ $F(6,12) = 32.8, p < 0.05$ ]. Bonferroni corrected *post hoc* comparisons indicated that for the 1.8 log MAR letter size, object frequency was significantly greater ( $p < 0.05$ ) for measurements in white noise

(triangles) compared to those measured in both filtered noise (squares) and in the absence of noise (circles). The *post hoc* comparisons also indicated a significant difference ( $p < 0.05$ ) between the object frequency measured in filtered noise (squares) and that measured in white noise (triangles) at the 1.2 log MAR letter size.

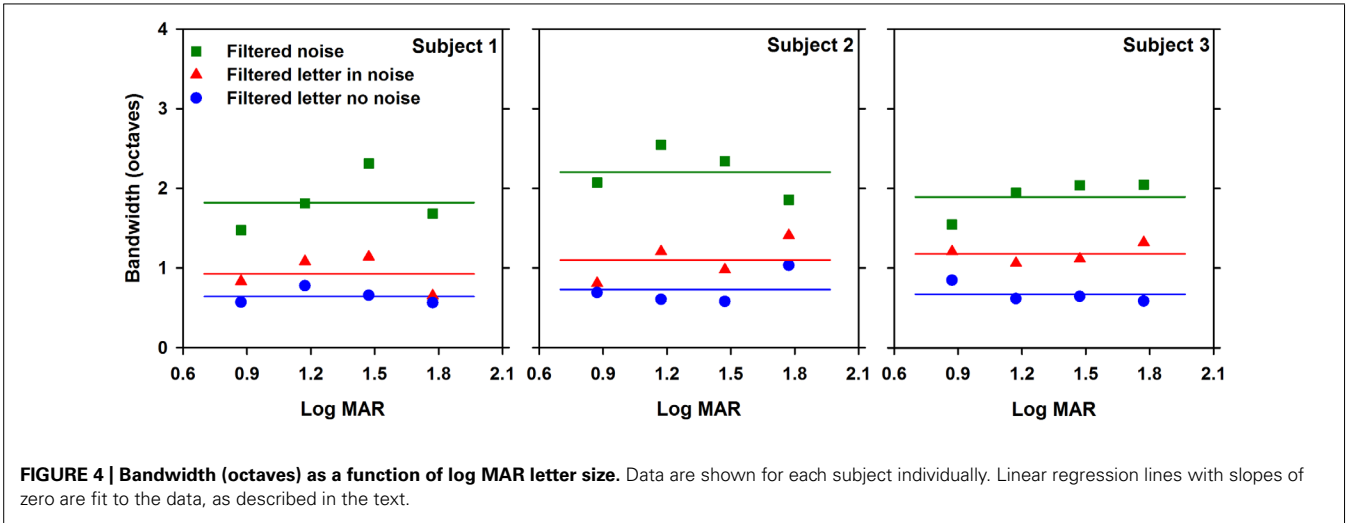
The bandwidth of useful object frequencies mediating letter contrast threshold was also assessed for each paradigm at each letter size, with the results for each subject shown in **Figure 4**. As in **Figure 3**, measurements are shown for letters in the absence of noise (circles), letters in the presence of white noise (triangles), and for letters in filtered noise (squares). The bandwidths for each paradigm were defined as the full width at half-max, as described above. The data were fit with linear regression lines of zero slope, since there was no effect of letter size on bandwidth, as discussed below. A two-way repeated measures ANOVA was performed to compare the bandwidths measured under the three paradigms at the four letter sizes. The ANOVA indicated significant main effects of paradigm [ $F(2,12) = 106.2, p < 0.05$ ], but not letter size [ $F(3,12) = 1.3, p > 0.05$ ]. Of note, the finding that bandwidth is approximately independent of letter size has been reported previously (Chung et al., 2002). Additionally, there was no significant interaction between paradigm and size [ $F(6,12) = 1.5, p > 0.05$ ]. The estimates of mean center frequency and bandwidth ( $\pm$ SEM) are listed in **Table 1** for each paradigm and letter size. Additionally, **Table 1** lists the mean ( $\pm$ SEM) contrast threshold for unfiltered letters in the absence and presence of noise and for unfiltered letters in white noise.

The relationship between the retinal spatial frequencies (cycles per degree; cpd) mediating letter identification and log MAR letter size is shown in **Figure 5**, which replots the data and best-fit curves of each subject shown in **Figure 3** in terms of cpd. This conversion is based on the following relationship (Alexander and McAnany, 2010):

$$F_r = \frac{12^* F_o}{\text{MAR}} \quad (2)$$

where  $F_r$  is retinal frequency in cpd,  $F_o$  is object frequency in cpl, and MAR is 1/5 of the letter size in arcmin.





The top *x*-axis in **Figure 5** indicates the nominal retinal frequencies corresponding to the log MAR values. This correspondence is based on the convention that 0 log MAR (20/20 Snellen equivalent) is equivalent to a retinal frequency of 30 cpd (Regan et al., 1981). This relationship assumes that an object frequency of 2.5 cpl (equivalent to the letter stroke width) governs performance at all sizes. The diagonal dashed line in **Figure 5** represents a one-to-one relationship between the derived center retinal frequency and the nominal retinal frequency, based on this assumption. If letter identification is governed by an object frequency range centered at 2.5 cpl for all sizes, then the nominal retinal frequency would be proportional to log MAR and the data would fall along the dashed line. It is apparent that none of the curves in **Figure 5** follow the dashed line. Rather, the data points for the two smallest

letter sizes tested (highest frequencies) fall near the dashed line, whereas the data points for the two largest letter sizes tested fall above the line. There is a divergence for the filtered letter in white noise function for the largest letter size (equivalent to 0.5 cpd), where the retinal frequency is slightly higher than the values at 1.0 and 2.0 cpd.

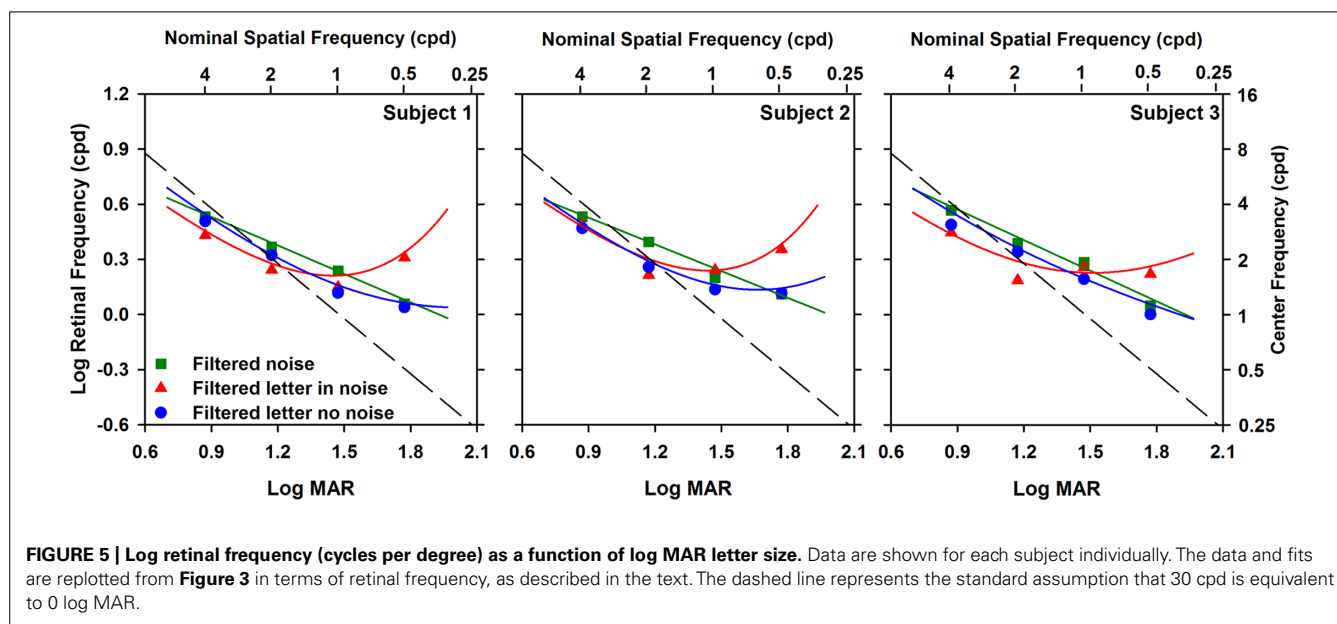
**DISCUSSION**

The purpose of the present study was to determine the effects of additive luminance noise on the object frequencies mediating letter contrast threshold across a range of letter sizes. The object frequency information mediating letter contrast threshold was assessed under three paradigms: (1) letters presented against a uniform adapting field; (2) letters presented in white luminance

**Table 1 | Center frequency, bandwidth, and contrast threshold for each letter size and paradigm.**

Paradigm	Size (log MAR)	Mean center frequency ± SEM (cpl)	Mean bandwidth ± SEM (octaves)	Log contrast threshold ± SEM (unfiltered letter/noise)
Filtered letter no noise	0.9	2.04 ± 0.05	0.70 ± 0.08	−1.64 ± 0.03
	1.2	2.61 ± 0.15	0.67 ± 0.06	−1.66 ± 0.07
	1.5	3.72 ± 0.20	0.63 ± 0.02	−1.89 ± 0.08
	1.8	5.95 ± 0.47	0.73 ± 0.15	−1.73 ± 0.02
Filtered letter in noise	0.9	1.89 ± 0.07	0.95 ± 0.13	−0.82 ± 0.07
	1.2	2.21 ± 0.08	1.12 ± 0.05	−0.71 ± 0.06
	1.5	4.35 ± 0.33	1.08 ± 0.05	−0.88 ± 0.03
	1.8	10.46 ± 0.92	1.13 ± 0.24	−0.95 ± 0.01
Filtered noise	0.9	2.31 ± 0.07	1.70 ± 0.19	−0.86 ± 0.01
	1.2	3.20 ± 0.06	2.10 ± 0.23	−0.81 ± 0.02
	1.5	4.60 ± 0.26	2.23 ± 0.10	−0.88 ± 0.08
	1.8	6.22 ± 0.29	1.86 ± 0.10	−1.08 ± 0.05





**FIGURE 5 | Log retinal frequency (cycles per degree) as a function of log MAR letter size.** Data are shown for each subject individually. The data and fits are replotted from **Figure 3** in terms of retinal frequency, as described in the text. The dashed line represents the standard assumption that 30 cpd is equivalent to 0 log MAR.

noise; (3) letters presented in filtered luminance noise (critical-band noise masking). There were similarities among the three paradigms in that the band of object frequencies mediating contrast threshold systematically increased with increasing letter size, consistent with previous reports (Alexander et al., 1994; Solomon and Pelli, 1994; Chung et al., 2002; Majaj et al., 2002; McAnany and Alexander, 2008; Oruc and Landy, 2009). However, the functions relating log object frequency to log MAR obtained under the three paradigms had three different shapes: For letters presented in white noise, the increase in log object frequency with increasing log MAR was strongly non-linear (exponential increase); For letters presented against a uniform field, the increase in log object frequency as log MAR increased was weakly non-linear; For letters presented in filtered noise, the increase in log object frequency with increasing log MAR was linear.

Despite the differences in the shapes of the functions relating object frequency and letter size, the absolute values of object frequency were approximately similar for letter sizes ranging from 0.9 to 1.5 log MAR. If the visual pathway mediating contrast threshold had changed from the magnocellular (MC) pathway to the parvocellular (PC) pathway due to the addition of noise, as proposed as a possibility in the Introduction, an increase in object frequency would be expected for letters presented in white noise. This was not observed. The explanation for why the object frequencies did not increase for small to medium sized letters due to the addition of noise is that the PC pathway likely mediated contrast threshold under all conditions. This explanation is based on the values of object frequency obtained previously under conditions biased toward the PC pathway (Alexander and McAnany, 2010), which closely match the values obtained in the presence and absence of noise in the present study. Additional work is needed to determine how the object frequency results of the present study would differ if measured under conditions that targeted the MC pathway.

A substantial increase in object frequency was observed for the largest letter size tested (equivalent to 1.8 log MAR) in white noise,

compared to the object frequencies used in the absence of noise or for letters in filtered noise. An increase in center object frequency due to the addition of white noise is also apparent in the data of Oruc and Landy (2009), but the consistency and significance of the change in their data is difficult to evaluate, as their focus was on determining whether object frequency depends on letter size in the presence of white noise. Oruc and Landy (2009) used critical-band noise masking to derive the object frequencies mediating letter identification for letters in the presence and absence of a white noise field. Despite differences in methodology, our results support their finding that the object frequencies mediating letter contrast threshold depend on letter size regardless of whether the letters are presented against a uniform field or in the presence of white noise.

A model has been proposed by Chung et al. (2002) to account for the shift in object frequency with letter size. This model suggests that the center frequency of the band of object frequencies mediating letter contrast threshold is jointly dependent on the letter sensitivity function (LSF; the object frequencies available in the letter) and the contrast sensitivity function (CSF; the relationship between contrast sensitivity and retinal frequency). This model accounts well for the change in object frequency with changes in letter size for letters of relatively small angular subtense (Chung et al., 2002). However, this model cannot account for the observed changes in center frequency for the large letter sizes used in the present study because the LSF and contrast sensitivity are both independent of letter size for letters greater than approximately 1.2 log MAR (McAnany and Alexander, 2006). Consequently, the model of Chung et al. (2002) would predict that center frequency is constant for large letters. Oruc and Landy (2009), who used white noise to flatten the CSF rather than using large letters, also reported that the center frequency changed with letter size.

At least two possible explanations could account for the large shift in object frequency due to the addition of white noise for

the 1.8 log MAR letter size. First, the noise fields in the present study were designed according to previous guidelines to be effectively white (Kukkonen et al., 1995), but the substantial power in the low frequency range for large check sizes may provide enhanced masking of the low object frequencies contained in the letters. Attenuation of the low object frequencies would force the observer to use higher object frequencies (i.e., the edges of the letters), accounting for the differences observed in the presence and absence of noise. Additional work with noise fields that have a constant check size for letters of different size is needed to evaluate this explanation. As an alternative explanation, the subjects may have employed a different strategy to perform the task in the presence of noise. Previous work has indicated that the addition of noise can affect the processing strategy used to perform the task (Allard and Cavanagh, 2011). Given that the differences in object frequency among the paradigms were not the same at all letter sizes, the effects of white noise are more likely attributable to strong masking of low object frequencies contained in the large letters, rather than a shift in processing strategy. However, additional work is needed to confirm this explanation.

In addition to the differences in center object frequency among the paradigms, there were also significant differences in bandwidth among the three paradigms. Bandwidth, averaged across subjects and letter size, was slightly greater in the presence of white noise (1.1 octaves) compare to measurements made in the absence of noise (0.7 octaves). The increase in bandwidth for letters in white noise tended to be similar for all letter sizes. Thus, for the largest letter tested, the center frequency of the band of object frequencies mediating letter contrast sensitivity shifted to higher frequencies but did not become broader. The bandwidth measured for letters in filtered noise (2.0 octaves, on average), is similar to that reported previously (Solomon and Pelli, 1994; Chung et al., 2002; Majaj et al., 2002). Channel switching or “off-frequency looking” would be expected to broaden the estimated bandwidth and may provide an explanation for the larger range of frequencies used in filtered noise. For example, in the presence of low-pass filtered noise, the subject could potentially use a channel with a higher peak frequency to avoid the low-pass noise (the opposite could occur in high-pass filtered noise). Previous results indicate that under some conditions, subjects do switch channels to improve the signal to noise ratio (Oruc and Landy, 2009).

In summary, we show that the addition of noise can affect the object frequency information mediating letter identification, particularly for large letters. For letters equivalent to 1.8 log MAR, the addition of noise had marked effects on the object frequency information mediating letter identification. This finding suggests that moderate to small letter sizes may be most appropriate for comparing letter contrast threshold in the presence and absence of noise because the assumption of noise-invariant processing largely holds.

## ACKNOWLEDGMENTS

This research was supported by National Institute of Health grants R00EY019510 (J. Jason McAnany) and P30EY001792 (departmental core grant), and an unrestricted departmental grant from Research to Prevent Blindness.

## REFERENCES

- Alexander, K. R., and McAnany, J. J. (2010). Determinants of contrast sensitivity for the tumbling E and Landolt C. *Optom. Vis. Sci.* 87, 28–36. doi: 10.1097/OPX.0b013e3181c61117
- Alexander, K. R., Xie, W., and Derlacki, D. J. (1994). Spatial-frequency characteristics of letter identification. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 11, 2375–2382. doi: 10.1364/JOSAA.11.002375
- Allard, R., and Cavanagh, P. (2011). Crowding in a detection task: external noise triggers change in processing strategy. *Vision Res.* 51, 408–416. doi: 10.1016/j.visres.2010.12.008
- Anderson, R. S., and Thibos, L. N. (1999). Sampling limits and critical bandwidth for letter discrimination in peripheral vision. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 16, 2334–2342. doi: 10.1364/JOSAA.16.002334
- Chung, S. T., Legge, G. E., and Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Res.* 42, 2137–2152. doi: 10.1016/S0042-6989(02)00092-5
- Garcia-Perez, M. A. (1998). Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Res.* 38, 1861–1881. doi: 10.1016/S0042-6989(97)00340-4
- Huang, C., Tao, L., Zhou, Y., and Lu, Z. L. (2007). Treated amblyopes remain deficient in spatial vision: a contrast sensitivity and external noise study. *Vision Res.* 47, 22–34. doi: 10.1016/j.visres.2006.09.015
- Kukkonen, H., Rovamo, J., and Nasanen, R. (1995). Masking potency and whiteness of noise at various noise check sizes. *Invest. Ophthalmol. Vis. Sci.* 36, 513–518.
- Levi, D. M., and Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *J. Neurosci.* 23, 2522–2526.
- Majaj, N. J., Pelli, D. G., Kurshan, P., and Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Res.* 42, 1165–1184. doi: 10.1016/S0042-6989(02)00045-7
- McAnany, J. J., Alexander, K. R., Genead, M. A., and Fishman, G. A. (2013). Equivalent intrinsic noise, sampling efficiency, and contrast sensitivity in patients with retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* 54, 3857–3862. doi: 10.1167/iov.13-11789
- McAnany, J. J., Alexander, K. R., Lim, J. I., and Shahidi, M. (2011). Object frequency characteristics of visual acuity. *Invest. Ophthalmol. Vis. Sci.* 52, 9534–9538. doi: 10.1167/iov.11-8426
- McAnany, J. J., and Alexander, K. R. (2006). Contrast sensitivity for letter optotypes vs. gratings under conditions biased toward parvocellular and magnocellular pathways. *Vision Res.* 46, 1574–1584. doi: 10.1016/j.visres.2005.08.019
- McAnany, J. J., and Alexander, K. R. (2008). Spatial frequencies used in Landolt C orientation judgments: relation to inferred magnocellular and parvocellular pathways. *Vision Res.* 48, 2615–2624. doi: 10.1016/j.visres.2008.02.012
- McAnany, J. J., and Alexander, K. R. (2009). Contrast thresholds in additive luminance noise: Effect of noise temporal characteristics. *Vision Res.* 49, 1389–1396. doi: 10.1016/j.visres.2008.10.012
- McAnany, J. J., and Alexander, K. R. (2010). Spatial contrast sensitivity in dynamic and static additive luminance noise. *Vision Res.* 50, 1957–1965. doi: 10.1016/j.visres.2010.07.006
- NAS-NRC. (1980). Recommended standard procedures for the clinical measurement and specification of visual acuity. Report of working group 39. *Adv. Ophthalmol.* 41, 103–148.
- Nordmann, J. P., Freeman, R. D., and Casanova, C. (1992). Contrast sensitivity in amblyopia: masking effects of noise. *Invest. Ophthalmol. Vis. Sci.* 33, 2975–2985.
- Oruc, I., and Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *J. Vis.* 9, 1–19. doi: 10.1167/9.9.4
- Parish, D. H., and Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Res.* 31, 1399–1415. doi: 10.1016/0042-6989(91)90060-I
- Pelli, E. (1990). Contrast in complex images. *J. Opt. Soc. Am. A* 7, 2032–2040. doi: 10.1364/JOSAA.7.002032
- Pelli, D. G., and Hoepner, J. A. (1989). Letters in noise: a visual test chart that “bypasses” the optics. *Noninvasive Assess. Vis. Syst.* 7, 103–106.
- Pelli, D. G., Levi, D. M., and Chung, S. T. (2004). Using visual noise to characterize amblyopic letter identification. *J. Vis.* 4, 904–920. doi: 10.1167/4.10.6

- Poder, E. (2003). Spatial-frequency spectra of printed characters and human visual perception. *Vision Res.* 43, 1507–1511. doi: 10.1016/S0042-6989(03)00216-5
- Regan, D., Raymond, J., Ginsburg, A. P., and Murray, T. J. (1981). Contrast sensitivity, visual acuity and the discrimination of Snellen letters in multiple sclerosis. *Brain* 104, 333–350. doi: 10.1093/brain/104.2.333
- Solomon, J. A., and Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature* 369, 395–397. doi: 10.1038/369395a0
- Xu, P., Lu, Z. L., Qiu, Z., and Zhou, Y. (2006). Identify mechanisms of amblyopia in Gabor orientation identification with external noise. *Vision Res.* 46, 3748–3760. doi: 10.1016/j.visres.2006.06.013
- Yates, J. T., Leys, M. J., Green, M., Huang, W., Charlton, J., Reed, J., et al. (1998). Parallel pathways, noise masking and glaucoma detection: behavioral and electrophysiological measures. *Doc. Ophthalmol.* 95, 283–299. doi: 10.1023/A:1001808407810

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 26 November 2013; accepted: 09 June 2014; published online: 03 July 2014.

Citation: Hall C, Wang S, Bhagat R and McAnany JJ (2014) Effect of luminance noise on the object frequencies mediating letter identification. *Front. Psychol.* 5:663. doi: 10.3389/fpsyg.2014.00663

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Hall, Wang, Bhagat and McAnany. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# To characterize contrast detection, noise should be extended, not localized

Rémy Allard<sup>1,2,3 \*</sup> and Jocelyn Faubert<sup>4,5</sup>

<sup>1</sup> INSERM, U968, Paris, France

<sup>2</sup> Institut de la Vision, Sorbonne Universités – University Pierre and Marie Curie, UMR\_S 968, Paris, France

<sup>3</sup> CNRS, UMR\_7210, Paris, France

<sup>4</sup> Visual Psychophysics and Perception Laboratory, Université de Montréal, Montréal, QC, Canada

<sup>5</sup> NSERC-Essilor Industrial Research Chair, Montréal, QC, Canada

## Edited by:

Denis Pelli, New York University, USA

## Reviewed by:

Jonathan W. Peirce, Nottingham  
University, UK

Niko Busch, Charité –  
Universitätsmedizin Berlin, Germany

## \*Correspondence:

Rémy Allard, INSERM, U968, Paris,  
F-75012, France  
e-mail: remy.allard@inserm.fr

Adding noise to a stimulus is useful to characterize visual processing. To avoid triggering a processing strategy shift between the processing in low and high noise, Allard and Cavanagh (2011) recommended using noise that is extended as a function of all dimensions such as space, time, frequency and orientation. Contrariwise, to avoid cross-channel suppression affecting contrast detection, Baker and Meese (2012) suggested using noise that is localized as a function of all dimensions, namely “0D noise,” which basically consists in randomly jittering the target contrast (and, for blank intervals or catch trials, jittering the contrast of an identical zero-contrast signal). Here we argue that contrast thresholds in extended noise are *not* contaminated by noise-induced cross-channel suppression because contrast gains affect signal and noise by the same proportion leaving the signal-to-noise ratio intact. We also review empirical findings showing that detecting a target in 0D noise involves a different processing strategy than detecting in absence of noise or in extended noise. Given that internal noise is extended as a function of all dimensions, we therefore recommend using external noise that is also extended as a function of all dimensions when assuming that the same processing strategy operates in low and high noise.

**Keywords:** external noise paradigm, extended noise, 0D noise, contrast detection, contrast discrimination

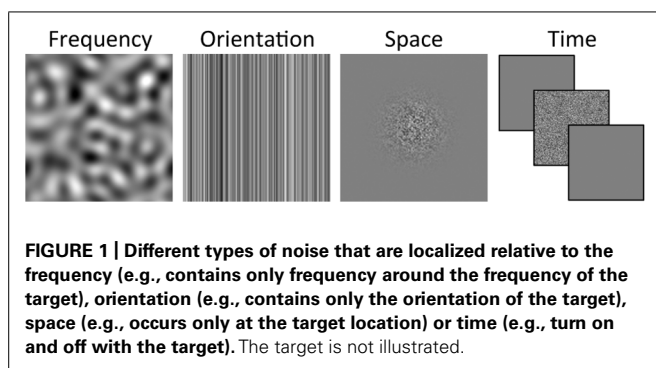
Noise can be used to characterize visual processing in noiseless conditions. For instance, contrast detection threshold in absence of noise is limited by both internal noise and the ability of detecting the signal embedded in noise, namely, calculation efficiency, which is inversely proportional to the smallest signal-to-(internal) noise ratio required to detect the signal. These two factors can be estimated by measuring contrast thresholds in low and high noise levels (Pelli, 1981; Pelli and Farell, 1999). In high noise, internal noise has negligible impact and the smallest signal-to-(external) noise ratio required to detect the signal (i.e., calculation efficiency in high noise) can be calculated given the contrast threshold in a given high external noise level. By assuming that the smallest signal-to-(internal) noise ratio required to detect the signal is the same as the measured smallest signal-to-(external) noise ratio required to detect the signal in high noise (i.e., assuming that the calculation efficiency in low noise is the same as the measured calculation efficiency in high noise), the relative impact of internal noise can be estimated, which is referred to as the internal equivalent noise. Thus, measuring contrast thresholds in low and high noise while assuming that the smallest signal-to-noise ratio required to detect the signal (i.e., calculation efficiency) is the same in low and high noise enables the measurement of factors limiting contrast thresholds in absence of noise, that is, internal equivalent noise and calculation efficiency.

Different types of noise can be used (Figure 1). Typically, noise turns on and off with the target (i.e., temporally localized) and appears at the target location (i.e., spatially localized)

or over slightly larger area. As a function of orientation and frequency, noise is often extended (e.g., white noise), that is, it has a wide spectral energy spectrum across orientations and frequencies. Nonetheless, it is not unusual to filter the noise to keep only a range of frequencies and orientations (or even only one orientation as in Figure 1).

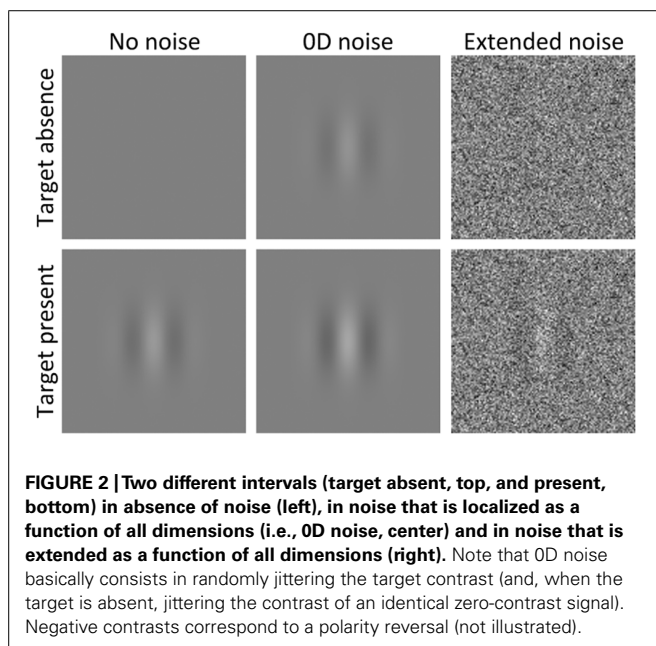
Typically, experimenters arbitrarily select one type of noise that is localized relative to some dimensions and extended relative to others, and usually implicitly assume that the smallest signal-to-(internal) noise ratio required to detect the signal is the same as the measured signal-to-(external) noise ratio. This assumption enables the use of external noise to characterize processing in noise free displays. However, some recent studies suggest that this assumption can be violated when using some types of noise. Allard and Cavanagh (2011) argued that the detection strategy is not always noise-invariant, as the most sensitive processes in one noise type may not be the most sensitive processes in another noise type. For instance, in noise that is spatially localized (appears only at the target location) and temporally extended (i.e., continuously present), the best detection strategy could consist in detecting a temporal variation of response within a given channel, but in noise that is spatially extended and temporally localized, the best strategy could rather consist in detecting a spatial variation. Two distinct strategies will likely have distinct calculation efficiencies. Thus, to assume that calculation efficiency in absence of noise is the same as in high noise, the same processing strategy must operate in absence and presence of noise. To avoid different processing





strategies operating in absence of noise and in high noise, external noise should match, as much as possible, the characteristics of internal noise (except for contrast). Internal noise is extended as a function of all dimensions as it occurs at all orientations and frequencies, it is not present only at the target location and it does not turn on and off with the target. Allard and Cavanagh (2011) therefore recommended using noise that is extended as a function of all dimensions, such as space, time, frequency and orientation (Figure 2 right).

Contrariwise, Baker and Meese (2012) criticized the use of extended noise due to cross-channel suppression as the response within one channel tends to suppress the responses of other nearby channels. For instance, noise extended as a function of orientation will introduce noise not only in the channels tuned to the signal orientation, but also to channels tuned to all other orientations, which may suppress the response within the relevant channels. To avoid cross-channel suppression affecting contrast detection threshold in high noise, Baker and Meese (2012) suggested to use noise that is localized as a function of all dimensions, which they refer to as “0D noise” (Figure 2 center), which basically consists in randomly jittering the target contrast (and, for blank intervals



or catch trials, jittering the contrast of an identical zero-contrast signal).

In sum, many experimenters use noise that is localized as a function of some dimensions and extended as a function of others, and implicitly assume that the calculation efficiency in low noise is the same as the measured calculation efficiency in high noise. However, given that internal noise is extended, this assumption may be violated in localized noise if different processing strategies operate in localized and extended noise, which would result in different processing strategies in low localized noise (i.e., when internal extended noise dominates) and high localized noise. It may also be violated in extended noise if noise-induced cross-channel suppression affects the measurement of the calculation efficiency in high noise. The objective of the present study was to determine which noise type (localized or extended) should be used to avoid violating the assumption that the calculation efficiency in low noise is the same as the measured calculation efficiency in high noise, which is necessary to characterize detection processing in noiseless conditions (e.g., measure internal equivalent noise and calculation efficiency). Note that because we see no reasons a priori why the noise should be localized as a function of some dimensions and extended as a function of others, the current article focused on the two extreme cases: noise extended or localized as a function of all dimensions. On the one hand, if noise-induced cross-channel suppression affects the measured calculation efficiency, this will likely be the case for any dimension. On the other hand, if noise should be analogous to internal noise to avoid triggering a processing strategy shift, then it should be extended as a function of all dimensions. The present article first investigated if adding 0D noise (i.e., noise localized as a function of all dimensions) triggers a shift in processing strategy and then investigated if noise-induced cross-channel suppression affects contrast thresholds in extended noise.

## NOISE-INVARIANT PROCESSING ASSUMPTION

0D noise has the advantage that it cannot induce cross-channel suppression because it contains energy only within the relevant channels. Thus, the usefulness of 0D noise to characterize the *detection* process depends on whether the same processing strategy operates in 0D noise as in absence of noise. In a two-interval forced-choice paradigm (2IFC), a contrast detection task consists in one interval containing the signal at a given contrast level and the other interval is blank (or contains an identical zero-contrast signal). For such a detection task, adding 0D noise consists in adding an independent contrast jitter to both intervals. As a result, a signal is presented in both intervals (e.g., Figure 2 center) and the task consists in discriminating the interval containing the highest contrast (while considering a contrast opposite to the signal as a negative contrast). In other words, a contrast detection task in 0D noise is processed as a contrast *discrimination* task (Allard and Faubert, 2013). Thus, if the processing strategies underlying contrast detection and discrimination tasks differ, then 0D noise could not be used to investigate the contrast detection process. Nevertheless, Baker and Meese (2013) argued that 0D noise can be used to characterize the *detection* process because, they claimed, a detection task is always processed as a discrimination task. In other words, they suggest that the processing strategy is the same for



contrast detection and discrimination tasks. If the same processing strategy operates for contrast detection and discrimination tasks, then 0D noise could indeed be used to characterize the detection process. On the other hand, if contrast detection, and discrimination tasks involves distinct processing strategies, this would disqualify the use of 0D noise to characterize the detection process and would provide further evidence that different processing strategies can operate in the absence of noise and presence of localized noise. Here, we review empirical findings suggesting that the detection strategy in noiseless conditions is not based on a discrimination strategy, as is the case for a detection task in 0D noise.

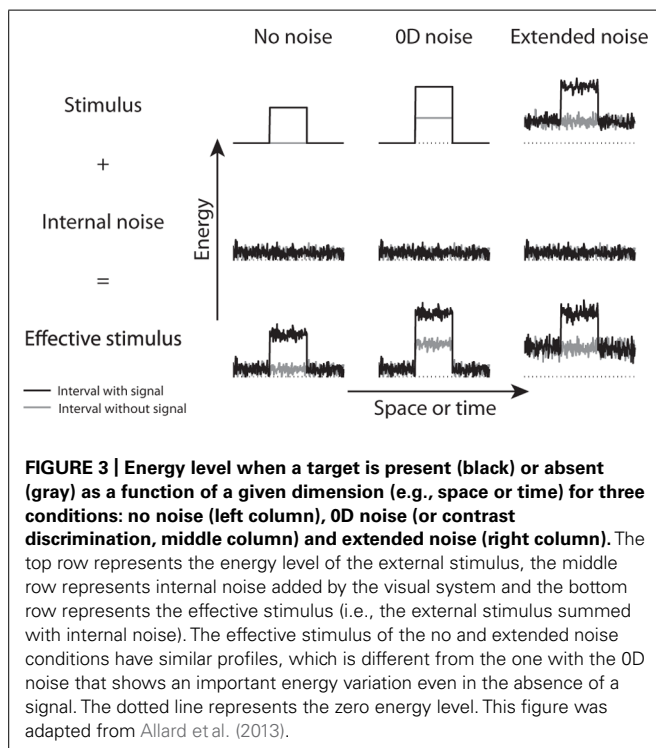
The contrast discrimination processing strategy is straightforward: compare two responses from the two intervals and report the one with the highest. Although such a strategy could also be used for contrast detection, it is not necessarily the case. As we suggested elsewhere (Allard and Cavanagh, 2011; Allard et al., 2013), an alternative detection strategy consists in determining if a pattern can be distinguished from the noisy background (Figure 3). According to this processing strategy, detection thresholds do not depend on the ability to discriminate two responses (as soon as the target is detected in one interval, the task is trivial), but simply on the ability to distinguish a pattern from the noisy background. Conversely, discrimination thresholds do not depend on the ability to distinguish a pattern from the noisy background (this is usually trivial because both stimuli are generally suprathreshold), but on the ability to estimate and compare two contrasts. According to Figure 3, the discrimination strategy would consist in comparing the energy levels in the central portion of the two curves (which could operate in any noise condition), whereas the detection strategy could also consist in

distinguishing a variation of energy relative to the background (which could not operate in 0D noise as the noise alone induces such a variation).

Allard and Cavanagh (2011) found empirical evidence that the detection strategy consists in distinguishing a pattern from the noisy background. They found that spatiotemporally localized noise (i.e., noise appearing only at the potential target spatiotemporal locations), which introduces energy easily distinguishable from the background (similar to the middle column in Figure 3), impaired the detection process and triggered a change in processing strategy: the processing strategy shifted from a detection strategy immune to crowding to a discrimination or recognition strategy that is sensitive to crowding. This processing strategy shift must be due to the spatiotemporal window of the noise and not to the noise *per se* because extended noise (i.e., background dynamic, white noise that is full screen and continuously displayed, Figure 3 right column) was not found to affect the detection strategy. If the detection strategy in absence of noise consists in discriminating two activity levels, then there is no reason why this strategy would change in localized, but not extended noise. Allard and Cavanagh (2011) therefore suggested that the detection strategy in noiseless displays consists in distinguishing a pattern from the background internal noise, not comparing activity levels (which would be the same in localized and extended noise).

A particularity of the detection process is that it can be facilitated by the superposition of a low-contrast pedestal. Indeed, contrast discrimination functions (i.e., contrast discrimination thresholds as a function of the pedestal contrast), which show a gradual shift from a detection task (zero contrast pedestal) to a contrast discrimination task (high contrast pedestal), typically show a dip when plotted in contrast units: low contrast pedestals facilitate contrast detection thresholds and high contrast pedestals impair contrast discrimination thresholds (for a review, see Solomon, 2009). Such a dipper function was observed in absence of noise and in extended noise (Pelli, 1981) and led Pelli to state that “The dip is of great theoretical interest because it indicates that the process of detection is similar with and without the noise mask.” (p. 123). Indeed, similar patterns of results with and without noise suggest common underlying processes. If the detection strategy in 0D noise were the same as in absence of noise, then we would also expect a similar dip with 0D noise. However, this is obviously not the case because the detection thresholds in 0D noise are close to the ideal performance (Allard and Faubert, 2013; Baker and Meese, 2013) so substantial facilitation is impossible. This absence of facilitation in 0D noise, and the facilitation in noiseless and extended noise suggest that the detection strategy in 0D noise (i.e., contrast discrimination strategy) differs with the detection strategy in noiseless or extended noise conditions.

To add further evidence that the detection strategy does not consist in discriminating contrasts, but rather consists in distinguishing a pattern from the noisy background, we conducted an additional experiment. We compared contrast thresholds obtained using two 2IFC procedures. In the detection condition, one interval contained the target and the other was blank. In the phase-discrimination condition, one interval contained the target and the other contained the same target but with a reversed contrast polarity (i.e., negative contrast). Thus, for a given target



contrast, the signed contrast difference between the two intervals in the phase-discrimination condition would be twice the one in the detection condition. If the processing strategy consists in comparing the signed contrast difference between the two intervals, then the contrast thresholds should be two times lower in the phase-discrimination condition. Indeed, the contrast difference between the two intervals would be the same in the two conditions when the contrast in the phase-discrimination condition would be half the contrast in the detection condition. Thus, we would expect the threshold in the detection condition to be twice the one in the phase-discrimination condition. Note that nonlinearities within the visual system could make this factor differ from 2. For instance, if the threshold depends on the energy difference (which is proportional to the squared contrast) between the two intervals (Raab et al., 1963) rather than the contrast difference, then we would expect the threshold in the detection condition to be  $\sqrt{2}$  times higher than in the phase-discrimination condition (energy doubles when increasing contrast by a factor of  $\sqrt{2}$ ). In any case, contrast thresholds would be non-negligibly lower in the phase-discrimination condition because, for a given target contrast, contrast difference (or energy difference) between the two intervals in the phase-discrimination condition would be twice the one in the detection condition. On the other hand, if the processing strategy consists in distinguishing a pattern from the noisy background, then the advantage in the phase-discrimination condition would only be due to the fact that two targets are presented compared to only one in the detection condition. The observer would have two chances instead of one to detect a target, so the observer would require a lower contrast level to obtain the same performance level. However, given that human observers have a sharp psychometric functions, performance drops rapidly when decreasing the target contrast so this advantage would only be of a factor of about 1.2 (Legge, 1984). Furthermore, this factor could be even less if the phase was not always discriminable when the target is detected because this would be a disadvantage in the phase-discrimination condition, but not in the detection condition.

## METHOD

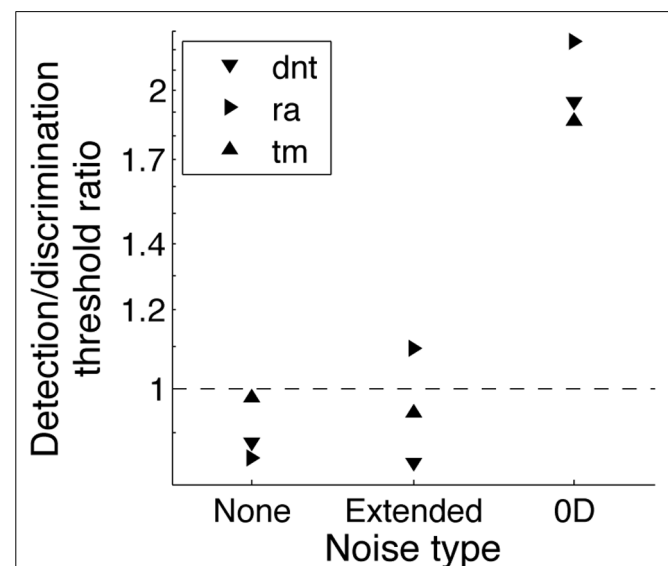
The target to detect was a vertically oriented Gabor with a spatial frequency of 0.7 cycles/degree and a standard deviation of the Gaussian window of  $0.5^\circ$ . The 0D noise contrast was 0.06 (standard deviation of the Gaussian distribution). The extended noise was binary with elements of  $2 \times 2$  pixels, resampled at 60 Hz and had a contrast of 0.32. The presentation time of each interval was 200 ms and the ISI was 500 ms. The contrast of the target was controlled by a 3-down-1-up staircase procedure (Levitt, 1971), which was interrupted after 12 inversions. To improve the luminance intensity resolution the Noisy-Bit method (Allard and Faubert, 2008) was implemented with the error of the green color gun inversely correlated with the error of the two other color guns, which made the 8-bit display perceptually equivalent to an analog display having a continuous luminance resolution. There were six block conditions (two tasks and three noise conditions, i.e., no noise, 0D noise and extended noise) that were performed three times each in a pseudorandom order. Contrast thresholds were estimated

as the geometric mean of the last eight inversions of the three blocks. Two naïve and one of the authors participated to the experiment.

## RESULTS AND DISCUSSION

In 0D noise (noise independently added in the two intervals), presenting a negative target instead of a blank interval improved threshold performance by a factor of about 2 (Figure 4). This was expected given that the detection strategy in 0D noise is a contrast discrimination strategy so contrast threshold depends on the contrast difference between the two intervals. In absence of noise and in extended noise, however, doubling the contrast difference between the two intervals (by switching from a detection to a phase-discrimination condition) did not result in a substantial threshold increase as the threshold ratio between these two conditions was close to 1. This suggests that the detection strategy in these conditions does not consist in discriminating contrasts between the two intervals while considering contrasts opposite to the target as negative contrasts, but rather in distinguishing a pattern from the noisy background.

In sum, the patterns of results observed for a contrast detection task in absence of noise were similar to the ones in extended noise and drastically different in 0D noise. Adding a low contrast pedestal substantially improves contrast thresholds in absence of noise and in extended noise, but not in 0D noise. Conversely, replacing the blank interval with a negative target substantially improved contrast thresholds in 0D noise, but not in absence of noise or in extended noise. This double dissociation between detection tasks in extended noise (whether internal or external) and in 0D noise suggests that they involve different processing strategies. Contrast detection thresholds in absence of noise or in extended noise reflect the ability to distinguish a pattern from



**FIGURE 4 |** Contrast thresholds for a 2IFC detection task relative to contrast thresholds for a 2IFC phase-discrimination task for three observers in three noise conditions: no noise, extended noise, and 0D noise.

the noisy background, not to discriminate contrasts in different intervals as in 0D noise.

### MODERATE 0D NOISE LEVEL

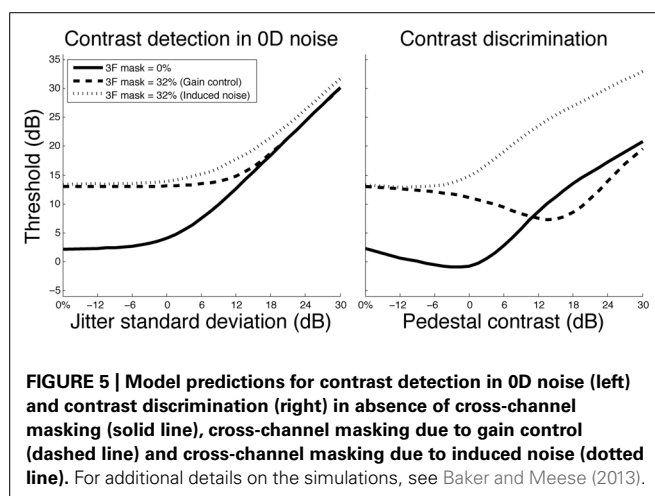
The section above suggests that measuring contrast detection thresholds in *high* 0D noise (i.e., when the impact of internal noise is negligible) cannot be used to characterize the detection process because such a stimulus is processed by a discrimination strategy that is distinct from the detection strategy operating in absence of noise. On the other hand, *low* 0D noise is also not useful to characterize the detection process as it has a negligible impact. Nonetheless, this does not rule out the possibility that moderate levels of 0D noise could be used to characterize the *detection* process. The present section will investigate if moderate levels of 0D noise can be used to characterize the detection process.

To empirically demonstrate the usefulness of 0D noise to characterize the detection process, Baker and Meese (2013) conducted an experiment in which they measured contrast detection thresholds as a function of noise contrast. Such a function usually shows, on a log-log plot, a smooth transition from a flat asymptote to a rising asymptote with a slope of 1 (e.g., **Figure 5** left). The flat asymptote can be evaluated by measuring detection threshold in absence of noise (or low noise). The rising asymptote in 0D noise can be evaluated by measuring contrast threshold in high 0D noise, but is known a priori as the task is trivial and the performance corresponds to the performance of an ideal observer (Allard and Faubert, 2013; Baker and Meese, 2013). Even though both asymptotes can be known without measuring any threshold in 0D noise, Baker and Meese (2013) showed that measuring contrast detection threshold as a function of 0D noise can be useful because different models predict different transitions between these two asymptotes. For instance, the gain control model would predict a smoother transition between the two asymptotes than the noise induce model (**Figure 5** left, see Baker and Meese, 2013, for model details). Given that a detection task is based on a detection strategy in low noise and a discrimination strategy in high 0D noise, the question is then to determine whether characterizing the transition between the two

asymptotes reveals properties of the detection or discrimination process.

Since 0D noise in a 2IFC procedure consists in contrast jittering both intervals independently, many trials in 0D noise are useless (even near threshold) because they can easily be discriminated, especially at high 0D noise contrasts. This leaves few trials in which the two stimuli have similar contrasts and the response is not trivial and will thereby depend on human factors, such as the ability to discriminate contrasts. Thus, if the 0D noise contrast is high enough to affect detection threshold, but not too high so that there is a non negligible proportion of trials in which both contrasts cannot be discriminated (i.e., around the transition between the two asymptotes), then contrast detection threshold in 0D noise would depend on contrast discrimination threshold. So if different models predict different contrast discrimination thresholds, they would also predict different contrast detection thresholds in moderate 0D noise. In other words, contrast detection threshold in moderate 0D noise would be an indirect, noisy measure of the contrast discrimination threshold. To illustrate this, we have replicated Baker and Meese's (2013) simulations for contrast detection threshold as function of 0D noise (**Figure 5** left) and ran the exact same simulations for a contrast discrimination task (i.e., the 0D noise was replaced by a pedestal, **Figure 5** right). Specifically, contrast thresholds as a function of external noise contrast (**Figure 5** left) and pedestal contrast (**Figure 5** right) were estimated by simulating trials using a standard detection model in which there was no masking (solid lines), the standard gain control model in which cross-channel masking is induced by suppression (dashed lines) and the noise-induced model in which cross-channel masking is induced by increasing internal noise (dotted lines). As illustrated in **Figure 5**, the two masking models, which affect contrast detection thresholds in absence of noise by the same proportion, predicted different contrast *discrimination* thresholds. This substantial contrast discrimination threshold difference directly explains the small contrast detection threshold difference in 0D noise. This shows that contrast detection threshold in moderate 0D noise is an indirect measure of the contrast discrimination process and that this experiment addresses the properties of the discrimination process, not the detection process.

Baker and Meese (2013) also showed that the two models predict different double pass consistencies. However, this property also directly depends on contrast discrimination thresholds. Indeed, the model predicting the higher contrast discrimination threshold will have the higher double pass consistency, as there will be more trials in which the two contrasts are discriminated. Given that the shape of the transition between the two asymptotes is directly related to contrast discrimination thresholds, we conclude that 0D noise could be used to investigate processing properties of the *discrimination* process, not detection process. In most cases, however, it would probably be more efficient to directly measure contrast discrimination thresholds. Nonetheless, even if there were some conditions in which measuring "detection" thresholds in 0D noise could be particularly useful to characterize the *discrimination* process, this would still not imply that measuring contrast detection thresholds in 0D noise can be useful to characterize the *detection* process.



## NOISE-INDUCED CROSS-CHANNEL SUPPRESSION

Baker and Meese (2012, 2013) argued that the use of white noise, which is extended as a function of frequency and orientation, is not suitable to measure internal equivalent noise because it induces cross-channel suppression affecting the measurement of contrast detection threshold in high noise thereby contaminating the measurement of the calculation efficiency. If the measurement of calculation efficiency in high noise were affected by noise-induced cross-channel suppression, then the assumption that the calculation efficiency in low noise is the same as the measured calculation efficiency in high noise would be compromised. Since contrast detection threshold in low noise depends on the internal equivalent noise and the calculation efficiency in low noise, not knowing the calculation efficiency in low noise would also compromise the estimation of the internal equivalent noise. The objective of the present section was to investigate if the assumption that the calculation efficiency in low noise is the same as the calculation efficiency in high noise is invalidated in extended noise due to noise-induced cross-channel suppression. Fortunately, we find that noise-induced cross-channel suppression does not affect contrast detection thresholds in high, extended noise for several reasons.

First, cross-channel suppression due to white noise seems weak. The strength of cross-channel suppression can be measured by asking the observer to match the contrast of a noise-free stimulus with the contrast of the same stimulus embedded in noise. Baker and Meese (2012) conducted such an experiment with 2D localized noise and their results were noisy: in some conditions, the noise had almost no impact on the perceived contrast and in others it affected threshold by a factor of about 2. This noise-induced suppression was not sufficient to explain the entire noise-induced threshold elevation of a factor of about 4. These results are inconsistent with previous findings showing that spatiotemporally extended white noise had no effect on perceived contrast (Pelli, 1981). To clarify this, we conducted our own contrast matching experiment and found that extended noise had no effect on perceived contrast (data not shown), which is consistent with Pelli's findings. Thus, determining if white noise affects perceived contrast (which would suggest some cross-channel suppression) remains an open question, but if it does, the effect would remain modest suggesting that noise-induced cross-channel suppression is weak at best.

Anyhow, determining if there is no or a weak noise-induced cross-channel suppression is irrelevant when measuring contrast thresholds in high noise. Any contrast gain affecting both the signal and the dominating noise source would have no impact on the signal-to-noise ratio and thereby would not affect contrast threshold. This is nicely illustrated by Baker and Meese's (2013) gain control model in which cross-channel suppression would affect contrast thresholds in low noise, but not in high noise (Figure 5, left). Indeed, when internal noise dominates (i.e., in low noise), a contrast gain occurring before the internal noise would affect the signal but not the dominating noise source and would therefore affect the signal-to-noise ratio. In high noise, however, the contrast gain would affect both the signal and the dominating noise source leaving the signal-to-noise ratio intact. Thus, even if noise reduced the effective contrast within the relevant channel

due to cross-channel suppression, this contrast reduction would not affect contrast thresholds.

Further evidence that noise-induced cross-channel suppression does not affect contrast thresholds in high noise comes from the fact that contrast thresholds in high noise are proportional to noise contrast (slope of 1 in log-log units as in Figure 5, left). This was first observed by Pelli (1981) and has been consistently replicated across many studies. To our knowledge, this fact has never been contradicted. This proportional relation between contrast threshold and noise contrast implies that contrast thresholds at distinct high noise contrasts result in the same signal-to-noise ratio and thereby the same measured calculation efficiency. The fact that the measured calculation efficiency in high noise is independent of the noise contrast even though extended noise induces more cross-channel suppression as its contrast is increased suggests that the measurement of the calculation efficiency is not affected by noise-induced cross-channel suppression. More generally, given that the signal-to-noise ratio required to detect the signal is independent of the noise contrast, there is no reason why this ratio would differ when the limiting noise source is internal only because the noise contrast is lower. We therefore conclude that noise-induced cross-channel suppression does not affect contrast thresholds in high noise and thereby does not compromise the assumption that the measured calculation efficiency in high noise is the same as the calculation efficiency in low noise and does not contaminate the measurement of calculation efficiency and internal equivalent noise limiting detection threshold in the absence of noise.

## CONCLUSION

Empirical findings suggest that different processing strategies operate for contrast detection in 0D noise compared to contrast detection in absence of noise and in extended noise. In 0D noise, the processing strategy consists in discriminating two contrasts, whereas in absence of noise (i.e., extended internal noise) and extended noise, the processing strategy consists in distinguishing a pattern from the noisy background. This suggests that different processing strategies operate in absence of noise and in 0D noise, which compromises the use of 0D noise to characterize the detection process operating in absence of noise (e.g., measure internal equivalent noise). Conversely, we found no evidence that the processing strategy differed in absence of noise and in extended noise, which suggests that extended noise could be used to characterize the detection process.

Baker and Meese (2012) suggested that high extended noise induces cross-channel suppression affecting contrast thresholds and thereby the measured calculation efficiency, which therefore could not be assumed to be the same as the calculation efficiency in absence of noise. However, this contrast reduction (if any) would not affect the contrast threshold as it would also affect the noise contrast thereby leaving intact the signal-to-noise ratio. This suggests that extended noise can be successfully used to characterize the detection process and measure internal equivalent noise.

In sum, the current study concludes that noise *extended* as a function of all dimensions can be used to characterize the contrast detection process, but noise *localized* as a function of all



dimensions cannot. Nevertheless, many experimenters use noise that is localized as a function of some dimensions and extended as a function of others. In principal, any property difference between internal and external noise could result in different detection strategies in low and high noise. On the other hand, a property difference between internal and external noise does not necessarily imply different detection strategies. For instance, a processing strategy could rely only on the central portion of a large stimulus and would therefore be independent of whether there is noise outside the stimulus region or not (i.e., spatially extended or localized noise, respectively; e.g., Allard and Faubert, 2014). Similarly, the processing strategy of a stimulus presented for a long duration would likely be independent of whether there is noise before and after the stimulus presentation or not (i.e., temporally extended or localized noise, respectively). Nonetheless, the detection strategy of a briefly presented, large stimulus could depend on whether the noise is temporally localized or extended (e.g., Allard and Faubert, 2014) and the detection strategy of a small stimulus presented for a long duration would likely depend on whether the noise is spatially localized or extended. Thus, using noise that is localized as a function of some dimensions raises doubts that the same detection strategy operates in low and high noise and thereby questions the assumption that the calculation efficiency in absence of noise is the same as the measured calculation efficiency in high noise. Given that internal noise is extended as a function of all dimensions, we therefore recommend using external noise that is also extended as a function of all dimensions when assuming that the same processing strategy operates in low and high noise.

## ACKNOWLEDGMENTS

This research was supported by NSERC discovery fund to Jocelyn Faubert and Essilor International.

## REFERENCES

- Allard, R., and Cavanagh, P. (2011). Crowding in a detection task: external noise triggers change in processing strategy. *Vision Res.* 51, 408–416. doi: 10.1016/j.visres.2010.12.008

- Allard, R., and Faubert, J. (2008). The noisy-bit method for digital displays: converting a 256 luminance resolution into a continuous resolution. *Behav. Res. Methods* 40, 735–743. doi: 10.3758/BRM.40.3.735
- Allard, R., and Faubert, J. (2013). Zero-dimensional noise is not suitable for characterizing processing properties of detection mechanisms. *J. Vis.* 13, 25. doi: 10.1167/13.10.25
- Allard, R., and Faubert, J. (2014). Motion processing: the most sensitive detectors differ in temporally localized and extended noise. *Front. Psychol.* 5:426. doi: 10.3389/fpsyg.2014.00426
- Allard, R., Renaud, J., Molinatti, S., and Faubert, J. (2013). Contrast sensitivity, healthy aging and noise. *Vision Res.* 92, 47–52. doi: 10.1016/j.visres.2013.09.004
- Baker, D. H., and Meese, T. S. (2012). Zero-dimensional noise: the best mask you never saw. *J. Vis.* 12, 20. doi: 10.1167/12.10.20
- Baker, D. H., and Meese, T. S. (2013). Regarding the benefit of zero-dimensional noise. *J. Vis.* 13, 26. doi: 10.1167/13.10.26
- Legge, G. E. (1984). Binocular contrast summation—I. Detection and discrimination. *Vision Res.* 24, 373–383. doi: 10.1016/0042-6989(84)90063-4
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49(Suppl. 2), 467+. doi: 10.1121/1.1912375
- Pelli, D. G. (1981). *The Effects of Visual Noise*. Ph.D. thesis, Cambridge University, Cambridge.
- Pelli, D. G., and Farell, B. (1999). Why use noise? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 16, 647–653. doi: 10.1364/JOSAA.16.000647
- Raab, D. H., Osman, E., and Rich, E. (1963). Effects of waveform correlation and signal duration on detection of noise bursts in continuous noise. *J. Acoust. Soc. Am.* 35, 1942–1946. doi: 10.1121/1.1918862
- Solomon, J. A. (2009). The history of dipper functions. *Atten. Percept. Psychophys.* 71, 435–443. doi: 10.3758/APP.71.3.435

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 January 2014; accepted: 26 June 2014; published online: 11 July 2014.

Citation: Allard R and Faubert J (2014) To characterize contrast detection, noise should be extended, not localized. *Front. Psychol.* 5:749. doi: 10.3389/fpsyg.2014.00749

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Allard and Faubert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Broadband noise masks suppress neural responses to narrowband stimuli

Daniel H. Baker\* and Greta Vilidaitė

Department of Psychology, University of York, York, UK

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Patrick J. Bennett, McMaster University, Canada

Denis Pelli, New York University, USA

## \*Correspondence:

Daniel H. Baker, Department of Psychology, University of York, Heslington, York YO10 5DD, UK  
e-mail: daniel.baker@york.ac.uk

White pixel noise is widely used to estimate the level of internal noise in a system by injecting external variance into the detecting mechanism. Recent work (Baker and Meese, 2012) has provided psychophysical evidence that such noise masks might also cause suppression that could invalidate estimates of internal noise. Here we measure neural population responses directly, using steady-state visual evoked potentials, elicited by target stimuli embedded in different mask types. Sinusoidal target gratings of 1 c/deg flickered at 5 Hz, and were shown in isolation, or with superimposed orthogonal grating masks or 2D white noise masks, flickering at 7 Hz. Compared with responses to a blank screen, the Fourier amplitude at the target frequency increased monotonically as a function of target contrast when no mask was present. Both orthogonal and white noise masks caused rightward shifts of the contrast response function, providing evidence of contrast gain control suppression. We also calculated within-observer amplitude variance across trials. This increased in proportion to the target response, implying signal-dependent (i.e., multiplicative) noise at the system level, the implications of which we discuss for behavioral tasks. This measure of variance was reduced by both mask types, consistent with the changes in mean target response. An alternative variety of noise, which we term zero-dimensional noise, involves trial-by-trial jittering of the target contrast. This type of noise produced no gain control suppression, and increased the amplitude variance across trials.

**Keywords:** noise masking, steady-state EEG, suppression, gain control, internal variability

## INTRODUCTION

Physical implementations of signal transduction systems suffer from degraded information transmission owing to internal noise. This is true both for electronic systems, such as amplifiers, and for biological sensory systems like the human visual system. It is of substantial interest to the study of basic perceptual processes (Kersten, 1984; Pelli, 1985; Legge et al., 1987; Gold et al., 1999; Allard and Faubert, 2006; Goris et al., 2008; Hess et al., 2008; Lu and Dosher, 2008; Baker and Meese, 2012), as well as clinical disorders (Pardhan et al., 1996; Levi and Klein, 2003; Pelli et al., 2004; Sperling et al., 2005; Xu et al., 2006; Huang et al., 2007; Milne, 2011), to be able to estimate the magnitude of this internal variability.

The standard method for estimating internal noise is to assess how task performance degrades in varying levels of external noise (Pelli, 1981; Lu and Dosher, 2008). The external noise will introduce variance into the detecting channel and raise thresholds for (e.g., reduce sensitivity to) target stimuli by decreasing the signal-to-noise ratio (see Appendix 1). The external noise level at which performance starts to become poorer is referred to as the “equivalent internal noise,” as it is the point at which the external noise is equal in magnitude to the internal noise. Various techniques exist for estimating this value, including fitting computational models (Lu and Dosher, 2008) and using Bayesian adaptive methods (Lesmes et al., 2006).

However, it has long been appreciated (Watson et al., 1997) that broadband white noise masks might have other effects on signal detection besides increasing within-mechanism variance.

There are several pieces of evidence that support a more complex account. Firstly, the slope of the psychometric function for contrast detection does not always become linear in noise (Klein and Levi, 2009; Baker and Meese, 2012), as would be predicted by Birdsall's theorem (Lasley and Cohn, 1981) for an individual nonlinear channel being swamped by external variance. Furthermore, the consistency of observer responses in noise across multiple passes through an identical trial sequence is lower for broadband noise than would be expected based on its masking potency (Burgess and Colborne, 1988; Lu and Dosher, 2008; Baker and Meese, 2012). Lastly, strong masking effects are observed even when the same sample of noise is used in both trial intervals (Watson et al., 1997; Beard and Ahumada, 1999; Baker and Meese, 2012); a result that would not occur for a noisy ideal observer limited only by variance.

What might be responsible for these deviations from the performance expected due to increased variance in the detecting channel? A plausible candidate is contrast gain control suppression (Heeger, 1992; Carandini and Heeger, 1994; Foley, 1994; Tolhurst and Heeger, 1997; Freeman et al., 2002; Sit et al., 2009; Carandini and Heeger, 2012) of the detecting mechanism by nearby mechanisms sensitive to other orientations and spatial frequencies also present in the noise mask. Several recent studies (Baker and Meese, 2012, 2013; Hansen and Hess, 2012) have provided behavioral evidence that supports this hypothesis. However, the possibility still remains that other processes, such as induced uncertainty or induced internal noise (Lu and Dosher, 2008), might be involved.

The present study used the steady-state visual evoked potential (SSVEP) technique (e.g., Tsai et al., 2012) to measure the neural response to contrast directly at the scalp. We show that broadband white noise masks have a powerful suppressive effect (see also Skoczinski and Norcia, 1998), very similar to that of narrowband orthogonal grating masks.

## MATERIALS AND METHODS

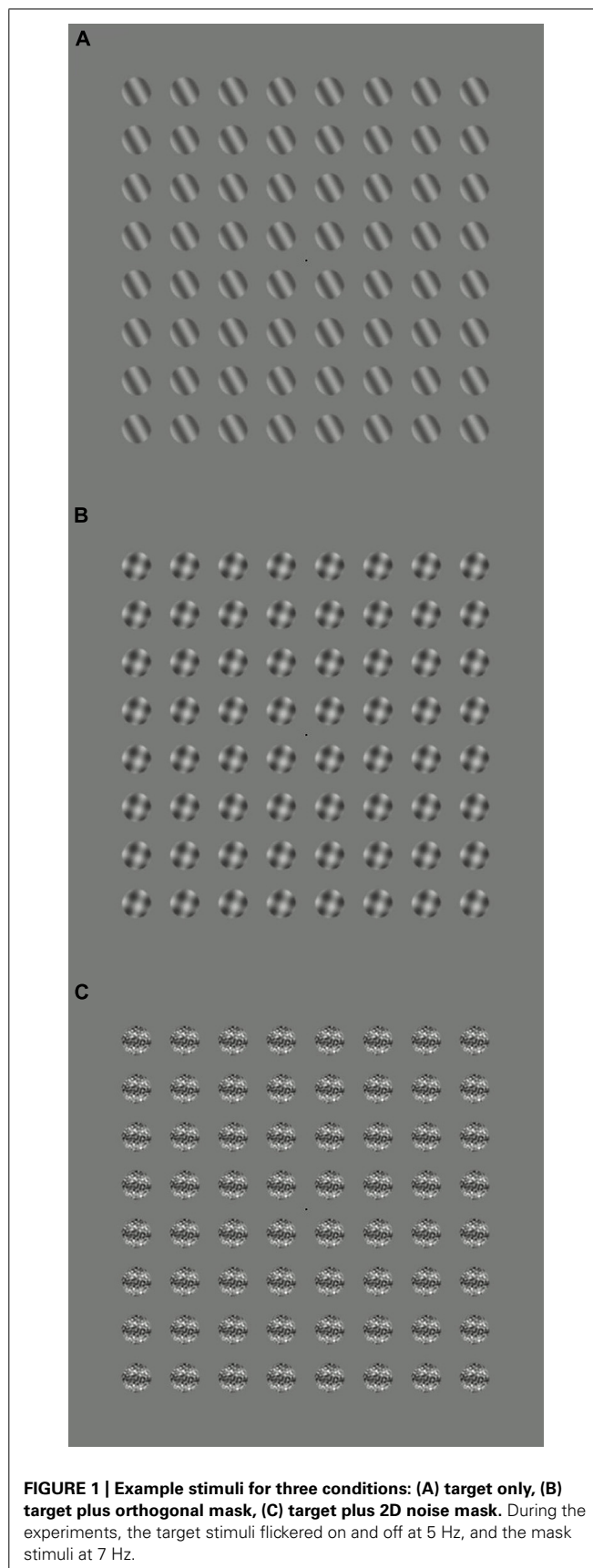
Stimuli were displayed on a gamma corrected Iiyama VisionMaster Pro 510 monitor using a Bits# stimulus generator (Cambridge Research Systems, Kent, UK). The monitor had a refresh rate of 75 Hz and a resolution of  $1024 \times 768$  pixels. When viewed from 57 cm, each degree of visual angle subtended 26 pixels on the display.

Target stimuli were patches of sine wave grating at 1 c/deg displayed at one of five Michelson contrasts (4, 8, 16, 32, or 64%), defined as  $C_{\%} = 100(L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$ , where  $L$  is luminance. Contrast is also expressed throughout in logarithmic (dB) units, where  $C_{\text{dB}} = 20 \log_{10}(C_{\%})$ . Stimuli increased and decreased in contrast (in linear units) according to a raised sine wave with a frequency of 5 Hz (on/off flicker), but did not reverse in phase. In the orthogonal mask condition, a second grating with a Michelson contrast of 32% was superimposed upon the target at right angles to it, flickering at 7 Hz. In the 2D noise condition, the mask was broadband white noise, low pass filtered at 5 c/deg, with an RMS contrast of 22%, and also flickering at 7 Hz. Note that the effect of the low pass filtering was to ensure that the majority of the noise power was not lost to very high spatial frequencies, where attenuation from the contrast sensitivity function is substantial. The noise remained white for more than two octaves above the target frequency. A new sample of noise was generated for each trial.

In the “0D” (zero dimensional) noise condition (Baker and Meese, 2012) the stimulus contrast was adjusted on a trial-by-trial basis. Contrasts were sampled from a normal distribution (in linear contrast units) with a standard deviation of 5.6% (15 dB) and added to the target contrast. When the total contrast was negative, the stimulus phase inverted. This meant that in practice the mean absolute contrasts of the lowest two target contrast levels increased to 5.6 and 8.4% in the 0D condition. The higher target contrasts were not materially affected by this phase inversion.

All stimuli were windowed by a circular raised cosine envelope with a blur width of 4 pixels ( $0.15^\circ$ ). They were tiled across the monitor in an  $8 \times 8$  grid (see **Figure 1**), and were displayed for trial durations of 11 s. To minimize adaptation, the orientation of the stimulus patches was randomized on every trial. There were five target contrast levels, and five stimulus configurations (no stimulus, target only, orthogonal mask, 2D noise mask, and 0D noise), which combined factorially to give 25 conditions. Observers completed five blocks, in which each condition was repeated twice (10 repetitions in total), taking around 1 h. Six adult observers completed the experiment; all had normal or optically corrected vision.

We recorded EEG signals at 64 electrode locations, distributed across the scalp according to the 10/20 system in a WaveGuard cap (ANT Neuro, Netherlands). We also recorded the vertical electrooculogram using self-adhesive electrodes placed above the



**FIGURE 1 | Example stimuli for three conditions: (A) target only, (B) target plus orthogonal mask, (C) target plus 2D noise mask.** During the experiments, the target stimuli flickered on and off at 5 Hz, and the mask stimuli at 7 Hz.

eyebrow and at the top of the cheek on the left side of the face. Signals were amplified and then digitized using a PC running the ASA software (ANT Neuro, Netherlands).

The data were imported into Matlab (Mathworks, MA, USA) and analyzed offline. We used average referencing to normalize all waveforms to the mean of all 64 electrodes (at each temporal sample). Each trial was split into eleven one second segments. The first 1 s was discarded to eliminate onset transients, and the remaining ten 1 s segments were Fourier transformed, with the phase and amplitudes recorded at both the target and mask frequencies (5 and 7 Hz). These ten observations were combined using coherent averaging to give a single measure of phase and amplitude for each trial at each electrode. We averaged across trials within each observer, and then calculated grand averages and standard errors across observers. The same procedure was used to average the signal variances.

## RESULTS

We first assessed activity at the stimulus frequencies across the electrode montage. We compared responses at 5 Hz between target absent trials, and trials on which the highest contrast target was shown in isolation. From **Figure 2A** it is clear that the strongest responses (largest colored circles) were observed at occipital electrodes. A similar pattern occurred for activity at 7 Hz when comparing target absent trials with the conditions in which either the orthogonal (**Figure 2B**) or 2D noise (**Figure 2C**) masks were shown along with the lowest contrast target. We therefore averaged waveforms across the two most active electrodes (*Oz* and *POz*) for the remaining analyses.

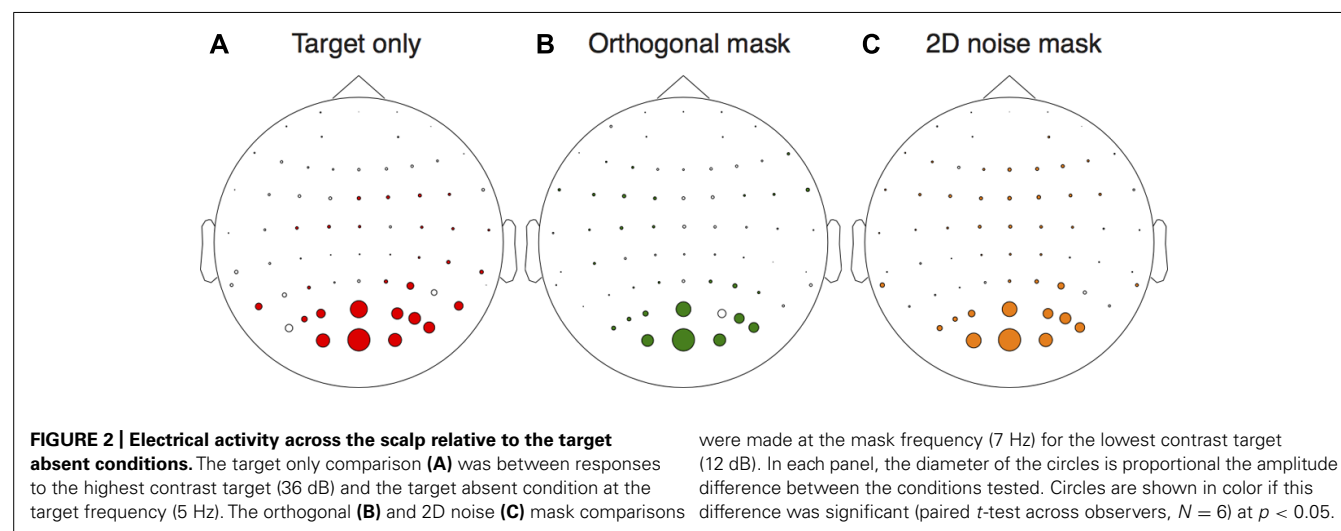
All observers produced responses that were monotonically increasing functions of target contrast. The average contrast response function to the target alone is shown by the red squares in **Figure 3A**. When a high contrast (30 dB) orthogonal mask was added at a higher temporal frequency (7 Hz), this shifted the contrast response function to the right (green triangles in **Figure 3A**). This is a classic contrast gain control effect, consistent with those reported in previous SSVEP (Brown et al., 1999; Busse et al., 2009; Tsai et al., 2012), fMRI (Brouwer and Heeger,

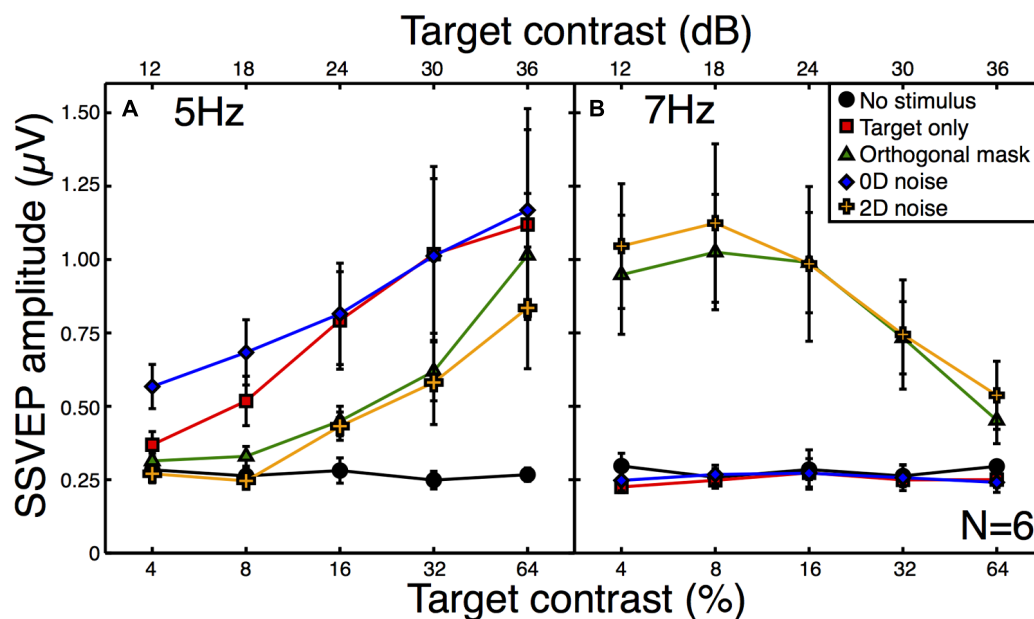
2011), and neuronal recordings (Morrone et al., 1982; Carandini and Heeger, 1994; Freeman et al., 2002; Busse et al., 2009; Sit et al., 2009).

The broadband white noise mask had a similar suppressive effect on the target response (orange crosses in **Figure 3A**), reducing the amplitude by a slightly greater amount than the orthogonal mask. This rightward shift of the contrast response function (also reported by Skoczenski and Norcia, 1998) is not predicted by standard noisy linear amplifier models of the noise masking process (Pelli, 1981; Lu and Dosher, 2008). There was also a strong response at the mask frequency to both of these masks (**Figure 3B**) which reduced as a function of target contrast. This illustrates the suppressive effects of the target onto the mask (Freeman et al., 2002; Busse et al., 2009; Brouwer and Heeger, 2011; Tsai et al., 2012) and confirms that inhibition occurs in both directions between the neural representations of the stimuli.

By way of comparison, we also measured responses in a 0D masking condition (blue symbols in **Figure 3A**). This involved jittering the stimulus contrast on a trial-by-trial basis. Although this manipulation might appear to make little sense for the single-trial observations of the SSVEP paradigm, it provides a useful comparison with psychophysical data. In 2AFC detection experiments, 0D noise is a very potent mask, raising thresholds by far more than 2D noise (Baker and Meese, 2012). However, it does this without reducing the mean neural response to the stimulus, as shown by the substantial overlap between red and blue functions in **Figure 3A**. The slightly greater response at the two lowest contrasts is easily understood when one considers that for weak target contrasts, large negative jitter values reverse the phase of the stimulus (see Materials and Methods). Since the SSVEP response is invariant with spatial phase it is the *absolute* contrast that determines the response, and this will be slightly higher than the nominal target contrast.

A second expectation of noise masks is that they increase the variance of neural responses across trials, because each unique noise sample will either increase or decrease activity in the detecting channel by a different amount (see Appendix 1). Note that





**FIGURE 3 | Steady-state visual evoked potential (SSVEP) amplitudes at the target (A) and mask (B) frequencies.** Amplitude values (with the phase component removed) were averaged across six observers, with error bars giving  $\pm 1$ SE of the mean. The target frequency was 5 Hz and the mask

frequency was 7 Hz. Note that for the “no stimulus” condition (black circles), the contrast is irrelevant (all contrasts were 0%), and the five points correspond to five separate repetitions of this baseline condition to illustrate the level of variability in our results.

the error bars in **Figure 3** are not a meaningful index of response variance, as they are calculated across (and not within) observers. To assess response variance, we calculated the trial-by-trial variance within observers for each condition, and then averaged these values across observers (**Figure 4A**). The variances clearly increase as a function of target contrast in all conditions. This is surprising, as it provides direct evidence of signal-dependent (i.e., multiplicative) noise within the visual system (Klein, 2006). The implications of this are discussed below.

One consequence of this signal-dependent noise is that the suppressive effect of the orthogonal mask also reduces the amplitude variance (green functions in **Figure 4A**). A similar reduction in variance is also produced by the 2D noise mask. This is rather worrying, as the aim of using external noise masks is typically to *increase* internal variance, not reduce it! Of course, a consequence of the frequency tagging used in the SSVEP procedure means that a variance increase at the signal frequency is unlikely, but a reduction is truly unexpected. The 0D noise produced an increase in variance at lower target contrasts, but no clear difference at higher target contrasts. This is presumably because the variance of the external noise mask was lower than the signal-dependent internal noise at these target contrast levels.

We also calculated the phase variance at the target frequency using circular statistics. The angular variance in radians was computed across epochs within each observer, and then averaged across observers to produce the plot in **Figure 4B**. High contrast stimuli produced responses that were strongly phase-locked, and so had low trial-by-trial variability. Low contrast stimuli lead to weaker phase locking, so the trial-by-trial variability was higher. The phase variance data in **Figure 4B** reveal a similar arrangement

of functions to the other figures, but inverted. This indicates a strong correspondence between signal amplitude and coherence (e.g., the inverse of variance).

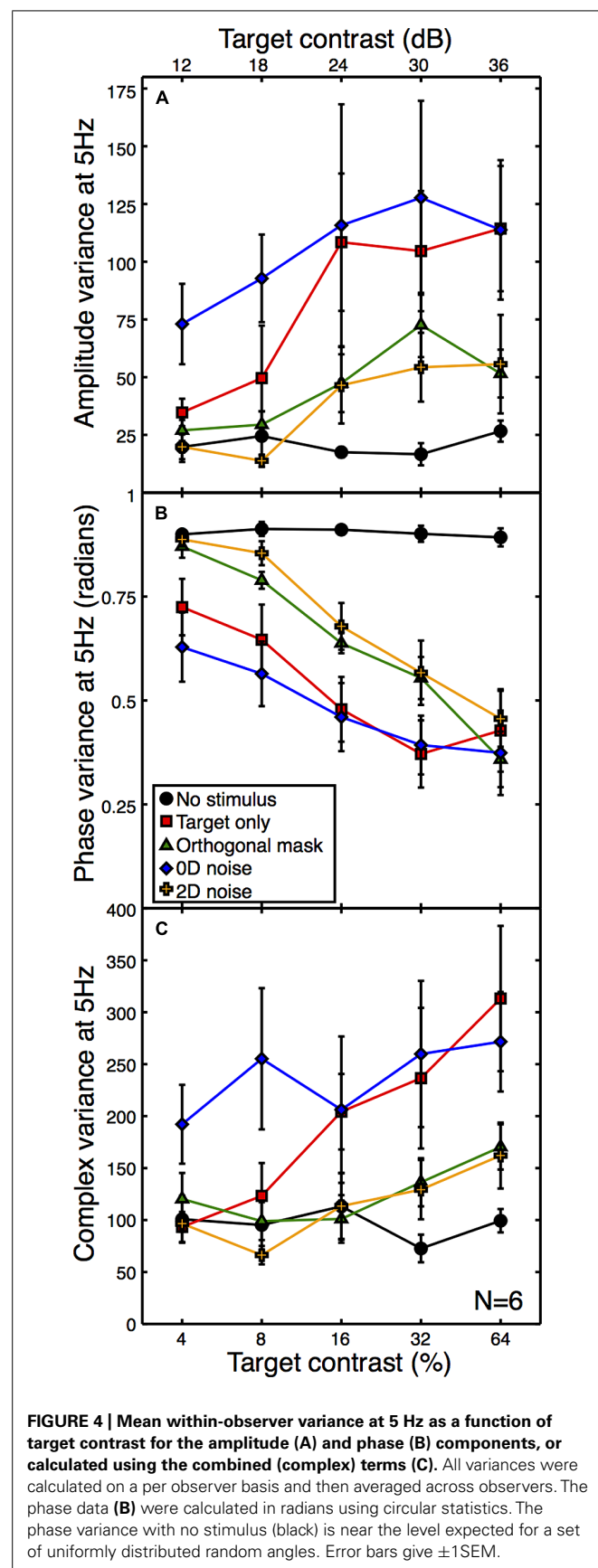
Note that as response amplitude increases, the amplitude variance increases but the phase variance decreases. It is therefore unlikely that the greater amplitude variance is a consequence of the phase locking of the SSVEP, as this would predict the opposite direction of effect to the one we report (e.g., amplitude variance would reduce for more coherent responses). However, we also calculated the variance of the complex Fourier components, which includes both phase and amplitude information. These are plotted in **Figure 4C**, and show a similar pattern to the data in **Figure 4A**, suggesting that the two individual variance measures are not confounded.

## DISCUSSION

We measured SSVEPs for patches of sine wave grating in the presence of different types of mask. The contrast response function was shifted rightward by orthogonal grating masks and by broadband noise masks. In addition, these mask types reduced the response variance, which we found to be proportional to the mean response. This pattern of responses suggests that broadband noise has a suppressive gain control effect on the neural response to the target. In comparison, a 0D noise manipulation where the signal strength was varied directly from trial to trial, did not reduce the mean response but did increase the response variance. This is the behavior expected of added external noise (see Appendix 1).

How might the steady-state responses correspond to an observer's perception, and the decisions they make in perceptual





tasks? We make the simplifying assumption that the VEP amplitude at the stimulus temporal frequency is proportional to the total neural population response to that stimulus, and that psychophysical decisions are based on the overall response, rather than the response of a subset of neurons. For contrast detection and discrimination experiments, this seems a reasonable assumption (e.g., Campbell and Maffei, 1970), though we note that it may not hold for more complex tasks (but see Ales et al., 2012). In addition, we made measurements at the occipital pole, which likely reflect activity in early visual areas. Later sources of internal noise could also influence an observer's decision in perceptual tasks. We note, however, that external noise is likely to have had its primary influence on neural responses by this stage.

Our results support previous misgivings about the ability of broadband noise to appropriately influence an observer's internal noise (Baker and Meese, 2012). Indeed, the observation that suppression reduces the multiplicative component of internal noise suggests that the problems may be more severe than previously suspected. Future noise masking studies would do well to limit the dimensionality and bandwidth of external noise stimuli as far as possible to mitigate the confounding effect of suppression. The 0D noise stimulus proposed by Baker and Meese (2012) might one way to achieve this aim in some experiments (e.g., Baker and Meese, 2014).

Interestingly, Skoczenski and Norcia (1998) have previously shown that broadband noise masks can shift the contrast response function to the right in both infants and adults. Although they acknowledge that contrast gain control may be responsible for their findings, they analyse their data based on the assumption that the external noise mask increases internal noise multiplicatively (e.g., see Lu and Dosher, 2008). The variance data shown in Figures 4A,C is inconsistent with this interpretation, as there is a clear reduction in variance when 2D noise masks are added (at least at the early stages of visual processing that contribute to occipital EEG signals). This speaks against the induced internal noise account of masking (see also Baker and Meese, 2013).

Steady-state VEP techniques are very well established, and have been used in countless studies. Given this ubiquity, we were surprised that previous reports of response-dependent noise were not forthcoming. This may be because the technique has often been used to address developmental (e.g., Skoczenski and Norcia, 1998; Brown et al., 1999) or clinical (e.g., Tsai et al., 2011) issues, rather than as a tool for basic research. We think that the combination of SSVEPs and computational modeling (see particularly Tsai et al., 2012) provides a valuable opportunity to investigate low-level sensory processes such as signal combination and suppression. In the following section, we use a modeling approach to show how the SSVEP data might be linked to psychophysical results.

#### RESPONSE-DEPENDENT NOISE: IMPLICATIONS FOR CONTRAST DISCRIMINATION

An unexpected finding was that response variance increases as a function of the mean response. Although this is well established at the level of individual neurons (Tolhurst et al., 1981, 1983), there is evidence that the dominant source of noise at a population level is effectively additive (Chen et al., 2006). In the psychophysics literature, there has been substantial debate over whether noise is



additive or multiplicative for behavioral tasks such as contrast discrimination (Kontsevich et al., 2002; Georgeson and Meese, 2006; Klein, 2006; Katkov et al., 2007). Pedestal masking effects (the Weber-like “handle” region of the dipper function) can either be obtained from a nonlinear transducer with additive noise (Legge and Foley, 1980), or a linear transducer with multiplicative noise (Pelli, 1985). Our results suggest that both may be present, since amplitude variance is response dependent (Figures 4A,C) and the contrast response function is nonlinear (Figure 3A). But which of these two features determines contrast discrimination behavior?

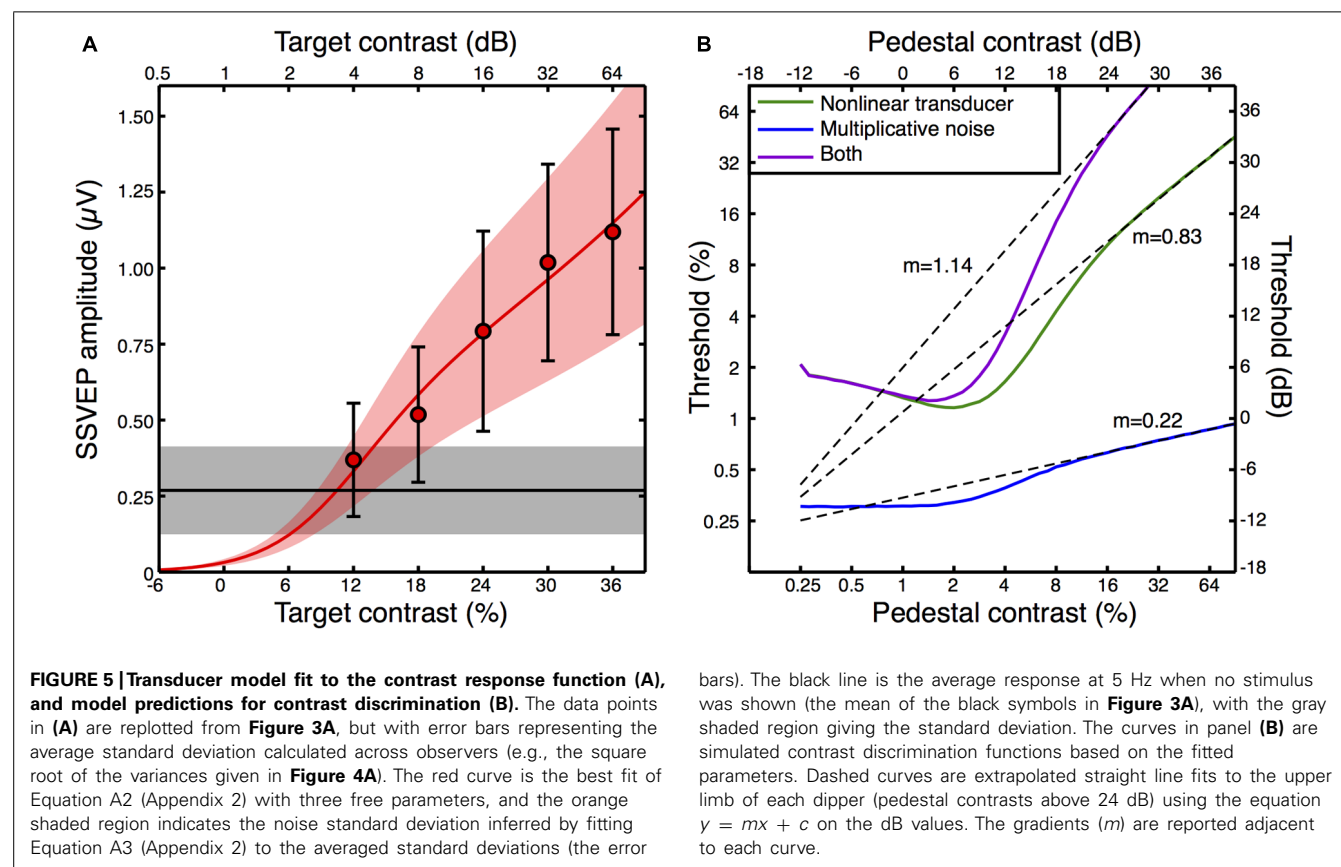
We fitted a transducer model to the amplitude and variance data from the average contrast response function (see Appendix 2 for details, and Figure 5A for the model fit). We then explored the predictions that three variants of this model made for psychophysical contrast discrimination experiments, as shown in Figure 5B. In the first variant, we set the multiplicative noise term (Equation A3 in Appendix 2) to zero. The green dipper function therefore shows the prediction based only on the transducer nonlinearity with additive noise. The second variant assumed a linear transducer ( $resp \propto C$ ) but with multiplicative noise proportional to the transduced contrast. This model variant, shown by the blue curve in Figure 5B, did not feature a dip. Typically facilitation is provided in such models by assuming that intrinsic uncertainty is reduced by the pedestal (Pelli, 1985). However we had no way to constrain such a model using our data set, and our exposition here focusses largely on the rising portion of the dipper. The slopes of the contrast discrimination functions were very different for these

two model variants, being 0.83 for the nonlinear transducer and 0.22 for the multiplicative noise model. Finally, we simulated a model that included both a transducer and multiplicative noise. The resulting dipper function (purple curve in Figure 5B) had a steeper handle, with a slope of 1.14.

The predicted dipper functions based on our steady state data appear plausible for the nonlinear transducer with additive noise, with a handle gradient somewhat steeper than the slope of  $\sim 0.6$  typically reported (Legge and Foley, 1980). When multiplicative noise is added, the handle becomes steeper still, yet even this value of  $>1$  is not inconsistent with previous reports using flickering stimuli similar to ours (Boynton and Foley, 1999). It therefore seems possible that previous attempts to estimate the underlying contrast response function based on psychophysical contrast discrimination data may be inaccurate if they neglect to include a multiplicative noise term. Historically, discrimination performance has been attributed to either a nonlinearity or multiplicative noise (Kontsevich et al., 2002; Georgeson and Meese, 2006; Klein, 2006; Katkov et al., 2007). To our knowledge, this is the first demonstration of how these two factors might combine.

## CONCLUSION

We have presented evidence that broadband noise masks have a suppressive gain control effect on neural responses to narrowband grating stimuli. This effect is similar to that obtained from orthogonal grating masks. Both mask types also reduce the amplitude variance, which is response dependent. We fitted a computational



model to the average contrast response function, and used this to infer the relative contribution of a nonlinear transducer and response-dependent noise for contrast discrimination. The modeling indicates that both features may be relevant to psychophysical contrast discrimination performance.

## ACKNOWLEDGMENTS

We are grateful to Alex Wade for helpful discussions relating to stimulus design and data analysis, and for comments on the manuscript.

## REFERENCES

- Ales, J. M., Farzin, F., Rossion, B., and Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *J. Vis.* 12, pii:18. doi: 10.1167/12.10.18
- Allard, R., and Faubert, J. (2006). Same calculation efficiency but different internal noise for luminance- and contrast-modulated stimuli detection. *J. Vis.* 6, 322–334. doi: 10.1167/6.4.3
- Baker, D. H., and Meese, T. S. (2012). Zero-dimensional noise: the best mask you never saw. *J. Vis.* 12, pii:20. doi: 10.1167/12.10.20
- Baker, D. H., and Meese, T. S. (2013). Regarding the benefit of zero-dimensional noise. *J. Vis.* 13, pii:26. doi: 10.1167/13.10.26
- Baker, D. H., and Meese, T. S. (2014). Measuring the spatial extent of texture pooling using reverse correlation. *Vision Res.* 97, 52–58. doi: 10.1016/j.visres.2014.02.004
- Beard, B. L., and Ahumada, A. J. Jr. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 16, 755–763. doi: 10.1364/JOSAA.16.000755
- Blakemore, C., and Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol.* 203, 237–260.
- Boynton, G. M., and Foley, J. M. (1999). Temporal sensitivity of human luminance pattern mechanisms determined by masking with temporally modulated stimuli. *Vision Res.* 39, 1641–1656. doi: 10.1016/S0042-6989(98)00199-0
- Brouwer, G. J., and Heeger, D. J. (2011). Cross-orientation suppression in human visual cortex. *J. Neurophysiol.* 106, 2108–2119. doi: 10.1152/jn.00540.2011
- Brown, R. J., Candy, T. R., and Norcia, A. M. (1999). Development of rivalry and dichoptic masking in human infants. *Invest. Ophthalmol. Vis. Sci.* 40, 3324–3333.
- Burgess, A. E., and Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *J. Opt. Soc. Am. A* 5, 617–627. doi: 10.1364/JOSAA.5.000617
- Busse, L., Wade, A. R., and Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64, 931–942. doi: 10.1016/j.neuron.2009.11.004
- Campbell, F. W., and Maffei, L. (1970). Electrophysiological evidence for the existence of orientation and size detectors in the human visual system. *J. Physiol.* 207, 635–652.
- Carandini, M., and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science* 264, 1333–1336. doi: 10.1126/science.8191289
- Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136
- Chen, Y., Geisler, W. S., and Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.* 9, 1412–1420. doi: 10.1038/nn1792
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. Opt. Image Sci. Vis.* 11, 1710–1719. doi: 10.1364/JOSAA.11.001710
- Freeman, T. C., Durand, S., Kiper, D. C., and Carandini, M. (2002). Suppression without inhibition in visual cortex. *Neuron* 35, 759–771. doi: 10.1016/S0896-6273(02)00819-X
- Georgeson, M. A., and Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out. *Vision Res.* 46, 4294–4303. doi: 10.1016/j.visres.2005.08.024
- Gold, J., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature* 402, 176–178. doi: 10.1038/46027
- Goris, R. L. T., Zaenen, P., and Wagemans, J. (2008). Some observations on contrast detection in noise. *J. Vis.* 8, 4.1–4.15. doi: 10.1167/8.9.4
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. (New York, NY: Wiley).
- Hansen, B. C., and Hess, R. F. (2012). On the effectiveness of noise masks: naturalistic vs. un-naturalistic image statistics. *Vision Res.* 60, 101–113. doi: 10.1016/j.visres.2012.03.017
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197. doi: 10.1017/S0952523800009640
- Hess, R. F., Baker, D. H., May, K. A., and Wang, J. (2008). On the decline of 1st and 2nd order sensitivity with eccentricity. *J. Vis.* 8, 19.1–19.12. doi: 10.1167/8.1.19
- Huang, C., Tao, L., Zhou, Y., and Lu, Z.-L. (2007). Treated amblyopes remain deficient in spatial vision: a contrast sensitivity and external noise study. *Vision Res.* 47, 22–34. doi: 10.1016/j.visres.2006.09.015
- Katkov, M., Tsodyks, M., and Sagi, D. (2007). Inverse modeling of human contrast response. *Vision Res.* 47, 2855–2867. doi: 10.1016/j.visres.2007.06.024
- Kersten, D. (1984). Spatial summation in visual noise. *Vision Res.* 24, 1977–1990. doi: 10.1016/0042-6989(84)90033-6
- Klein, S. A. (2006). Separating transducer non-linearities and multiplicative noise in contrast discrimination. *Vision Res.* 46, 4279–4293. doi: 10.1016/j.visres.2006.03.032
- Klein, S. A., and Levi, D. M. (2009). Stochastic model for detection of signals in noise. *J. Opt. Soc. Am. Opt. Image Sci. Vis.* 26, B110–B126. doi: 10.1364/JOSAA.26.00B110
- Kontsevich, L. L., Chen, C.-C., and Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Res.* 42, 1771–1784. doi: 10.1016/S0042-6989(02)00091-3
- Lasley, D. J., and Cohn, T. E. (1981). Why luminance discrimination may be better than detection. *Vision Res.* 21, 273–278. doi: 10.1016/0042-6989(81)90121-8
- Legge, G. E., and Foley, J. M. (1980). Contrast masking in human vision. *J. Opt. Soc. Am.* 70, 1458–1471. doi: 10.1364/JOSA.70.001458
- Legge, G. E., Kersten, D., and Burgess, A. E. (1987). Contrast discrimination in noise. *J. Opt. Soc. Am. A* 4, 391–404. doi: 10.1364/JOSAA.4.000391
- Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., and Doshier, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: the quick TvC method. *Vision Res.* 46, 3160–3176. doi: 10.1016/j.visres.2006.04.022
- Levi, D. M., and Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *J. Neurosci.* 23, 2522–2526.
- Lu, Z.-L., and Doshier, B. A. (2008). Characterizing observers using external noise and observer models: assessing internal representations with external noise. *Psychol. Rev.* 115, 44–82. doi: 10.1037/0033-295X.115.1.44
- Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: evidence from single-trial analysis of evoked EEG. *Front. Psychol.* 2:51. doi: 10.3389/fpsyg.2011.00051
- Morrone, M. C., Burr, D. C., and Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. neurophysiological evidence. *Proc. R. Soc. Lond. B Biol. Sci.* 216, 335–354. doi: 10.1098/rspb.1982.0078
- Pardhan, S., Gilchrist, J., Elliott, D. B., and Beh, G. K. (1996). A comparison of sampling efficiency and internal noise level in young and old subjects. *Vision Res.* 36, 1641–1648. doi: 10.1016/0042-6989(95)00214-6
- Pelli, D. G. (1981). *The Effects of Visual Noise*. Unpublished Ph.D. thesis, Cambridge University, Cambridge.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *J. Opt. Soc. Am. A* 2, 1508–1532. doi: 10.1364/JOSAA.2.001508
- Pelli, D. G., Levi, D. M., and Chung, S. T. L. (2004). Using visual noise to characterize amblyopic letter identification. *J. Vis.* 4, 904–920. doi: 10.1167/4.10.6
- Sit, Y. F., Chen, Y., Geisler, W. S., Miikkulainen, R., and Seidemann, E. (2009). Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron* 64, 943–956. doi: 10.1016/j.neuron.2009.08.041
- Skoczenski, A. M., and Norcia, A. M. (1998). Neural noise limitations on infant visual sensitivity. *Nature* 391, 697–700. doi: 10.1038/35630
- Sperling, A. J., Lu, Z.-L., Manis, F. R., and Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nat. Neurosci.* 8, 862–863. doi: 10.1038/nn1474
- Tolhurst, D. J., and Heeger, D. J. (1997). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Vis. Neurosci.* 14, 293–309. doi: 10.1017/S0952523800011433

- Tolhurst, D. J., Movshon, J. A., and Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 23, 775–785. doi: 10.1016/0042-6989(83)90200-6
- Tolhurst, D. J., Movshon, J. A., and Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Exp. Brain Res. Exp.* 41, 414–419.
- Tsai, J. J., Norcia, A. M., Ales, J. M., and Wade, A. R. (2011). Contrast gain control abnormalities in idiopathic generalized epilepsy. *Ann. Neurol.* 70, 574–582. doi: 10.1002/ana.22462
- Tsai, J. J., Wade, A. R., and Norcia, A. M. (2012). Dynamics of normalization underlying masking in human visual cortex. *J. Neurosci.* 32, 2783–2789. doi: 10.1523/JNEUROSCI.4485-11.2012
- Watson, A. B., Borthwick, R., and Taylor, M. (1997). Image quality and entropy masking. *SPIE Proc.* 3016, 1–11. doi: 10.1117/12.274501
- Xu, P., Lu, Z.-L., Qiu, Z., and Zhou, Y. (2006). Identify mechanisms of amblyopia in Gabor orientation identification with external noise. *Vision Res* 46, 3748–3760. doi: 10.1016/j.visres.2006.06.013
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 30 November 2013; accepted: 29 June 2014; published online: 15 July 2014.
- Citation: Baker DH and Vilidaitė G (2014) Broadband noise masks suppress neural responses to narrowband stimuli. *Front. Psychol.* 5:763. doi: 10.3389/fpsyg.2014.00763
- This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.
- Copyright © 2014 Baker and Vilidaitė. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## APPENDIX 1 – ASSUMPTIONS ABOUT NOISE MASKING

A primary tenet of signal detection theory (Green and Swets, 1966) is that performance on a task is determined by the magnitude of a scalar internal response variable. This response is determined by the amplitude of an external signal, and internal variability (noise) within the system. Within the framework of psychophysical “channels” sensitive to a limited range of orientations and spatial frequencies (Blakemore and Campbell, 1969), the internal response in a detection task will be determined by the energy falling within the pass-band of a linear filter, plus internal noise.

When external noise is added to a stimulus, some of the noise power will also fall within the pass-band of the detecting channel. On some presentations, this will increase the mechanism response, whereas on other presentations it will decrease it. Thus, the external noise will introduce variance into the internal response that governs decisions. A formal expression of this process is given by,

$$resp = (C_{test} + \sigma_{ext}R_1)^\gamma + \sigma_{int}R_2, \quad (A1)$$

where  $C_{test}$  is the target contrast,  $\sigma_{ext}$  is the standard deviation of external noise falling within the pass-band of the detecting channel,  $\gamma$  typically has a value of around 2, and  $\sigma_{int}$  is the standard deviation of internal noise (notation from Klein and Levi, 2009). The terms  $R_1$  and  $R_2$  denote samples from a Gaussian random number generator that are drawn independently on each trial of an experiment. In this model, the observer bases their decision *only* on the scalar response value; any noise power outside of the detecting channel is ignored.

A number of clear predictions follow from this model that can be tested empirically (e.g., Klein and Levi, 2009; Baker and Meese, 2012). In addition, several elaborations have been proposed that include features such as induced internal noise (Burgess and Colborne, 1988; Lu and Dosher, 2008), gain control suppression (Watson et al., 1997; Baker and Meese, 2012, 2013; Hansen and Hess, 2012) and uncertainty when selecting from multiple channels (Pelli, 1985).

Some studies have designed noise stimuli intended to target a particular stage of processing (Allard and Faubert, 2006; Baker and Meese, 2014), with the aim of increasing the proportion of the noise power that falls within the pass-band of the appropriate detecting mechanism. However, we note that any such manipulations can only influence decision behavior by changing the magnitude of the internal response variable, and so are equivalent to increasing the effective level of the external noise.

## APPENDIX 2 – DETAILS OF CONTRAST TRANSDUCTION MODELS

We fitted a standard transducer nonlinearity (Legge and Foley, 1980) to the target-only contrast response function. The nonlinearity is given by,

$$resp = R_{max} \frac{C^p}{Z + C^q}, \quad (A2)$$

where  $C$  is contrast,  $p$  and  $q$  are exponents,  $Z$  determines the gain, and  $R_{max}$  is a scaling parameter. To reduce the number of free parameters, we fixed  $q$  at the standard value of 2 (Legge and Foley, 1980). We minimized the root-mean-square (RMS) error between the mean amplitude and the model response, and obtained parameter estimates of  $p = 2.24$ ,  $Z = 12.13$  and  $R_{max} = 4.22$ . The fit is shown by the red curve in **Figure 5A**.

We then estimated a scaling parameter for multiplicative noise. Since the noise is clearly response-dependent rather than signal dependent (e.g., in **Figures 4A–C** the variances are reduced by the masks), we made the noise proportional to the transducer response,

$$noise = GN \times resp, \quad (A3)$$

where  $G$  represents samples of zero-mean Gaussian noise, with standard deviation determined by the output of Equation A2 ( $resp$ ) and a scaling factor,  $N$ . We estimated that  $N = 0.35$  by finding the value that best described the standard deviations of the responses, shown by the error bars in **Figure 5A** (note that these error bars are the square root of the mean within-observer amplitude variance values given by the red function in **Figure 4A**, and are not the same as the between-observer standard errors in **Figure 3A**). The model noise standard deviation is given by the orange shaded region in **Figure 5A**.

The fitted model parameters were then used to simulate thresholds for contrast discrimination (dipper) functions. To derive predictions at low contrasts, we also required an estimate of fixed (additive) noise. This was obtained from the variance in the target absent condition of the experiment (black symbols in **Figure 4A**). The horizontal black line and gray shaded rectangle in **Figure 5A** show the mean and standard deviation of the 5 Hz amplitude in this condition. We simulated a method of constant stimuli contrast discrimination experiment using the above equations and parameters, with 100,000 stochastic trials per target contrast level. Thresholds were obtained by fitting cumulative Gaussian functions to the simulated data.

The above simulations make several assumptions that may or may not be valid. Least plausible is perhaps our decision to use the variance at 5 Hz in the signal absent condition as an estimate of fixed (additive) internal noise. We think it highly unlikely that the 5 Hz variance when no stimulus is present represents the activity of neurons that subsequently respond to the stimulus, at least in any straightforward way. Many unrelated sources of variance will contribute to this figure, including equipment noise, electromagnetic interference, and spontaneous neural oscillations, so the noise baseline is likely to be an overestimate of the true variance. However, the additive noise term only influences detection and low-contrast discrimination performance (the leftmost parts of the dipper) and has little effect on the slope of the dipper handle. We repeated our simulations for several alternative additive noise levels, and found that the dipper handle gradients remained remarkably constant over a wide range.



# Developmental mechanisms underlying improved contrast thresholds for discriminations of orientation signals embedded in noise

Seong Taek Jeon<sup>1\*</sup>, Daphne Maurer<sup>2</sup> and Terri L. Lewis<sup>2</sup>

<sup>1</sup> Department of Vision Sciences, Institute for Applied Health Research, Glasgow Caledonian University, Glasgow, UK

<sup>2</sup> Visual Development Laboratory, Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Oliver Braddick, University of Oxford, UK

Helle K. Falkenberg, Buskerud and Vestfold University College, Norway

## \*Correspondence:

Seong Taek Jeon, Department of Vision Sciences, Institute for Applied Health Research, Glasgow Caledonian University, Cowcaddens Rd., Glasgow G4 0BA, UK  
e-mail: sje1@gcu.ac.uk

We combined an external noise paradigm with an efficient procedure for obtaining contrast thresholds (Lesmes et al., 2006) in order to model developmental changes in the effect of noise on contrast discrimination during childhood. Specifically, we measured the contrast thresholds of 5-, 7-, 9-year-olds and adults ( $n = 20/\text{age}$ ) in a two alternative forced-choice orientation discrimination task over a wide range of external noise levels and at three levels of accuracy. Overall, as age increased, contrast thresholds decreased over the entire range of external noise levels tested. The decrease was greatest between 5 and 7 years of age. The reduction in threshold after age 5 was greater in the high than the low external noise region, a pattern implying greater tolerance of the irrelevant background noise as children became older. To model the mechanisms underlying these developmental changes in terms of internal noise components, we adapted the original perceptual template model (Lu and Doshier, 1998) and normalized the magnitude of performance changes against the performance of 5-year-olds. The resulting model provided an excellent fit ( $r^2 = 0.985$ ) to the contrast thresholds at multiple levels of accuracy (60, 75, and 90%) across a wide range of external noise levels. The improvements in contrast thresholds with age were best modeled by a combination of reductions in internal additive noise, reductions in internal multiplicative noise, and improvements in excluding external noise by template retuning. In line with the data, the improvement was greatest between 5 and 7 years of age, accompanied by a 39% reduction in additive noise, 71% reduction in multiplicative noise, and 45% improvement in external noise exclusion. The modeled improvements likely reflect developmental changes at the cortical level, rather than changes in front-end structural properties (Kiorpes et al., 2003).

**Keywords:** vision, contrast thresholds, internal noise, development, psychophysics

## INTRODUCTION

Many aspects of basic vision improve rapidly over the first few years of life. For example, visual acuity, whether measured using visually-evoked potentials or preferential looking, improves rapidly between birth and 6 months of age and then continues to improve gradually until about 6 years of age (Norcia and Tyler, 1985; Chandna, 1991; Neu and Sireteanu, 1997). Front-end changes make a substantial contribution to the early improvements in basic visual abilities (e.g., Yuodelis and Hendrickson, 1986; Banks and Bennett, 1988). However, the development of cortical pathways appears to also contribute to the changes both during and after infancy (Banks and Bennett, 1988; Toga et al., 2006; Braddick and Atkinson, 2011).

The human brain is a complex system that consists of hundreds of anatomical structures and billions of neurons exchanging, at any given time, thousands of electrical and chemical signals through synapses connecting neurons in both nearby and remote parts of the brain. Like any other machinery, be it artificial or biological, neurons are not ideal transmitters of information.

For example, identical signals from neighboring neurons do not elicit identical responses in the receiving neurons each time they are produced. This variability is observed even when external conditions, such as the sensory input or task goal, are kept as constant as possible (Cohn and Lasley, 1986; Faisal et al., 2008).

How is this fluctuation manifested at the level of visual behavior? As an illustration, a visual object may cause activation of neurons responsible for signaling its particular visual properties embedded in background activation of some neurons irrelevant to those properties. The spontaneous activation of these other neurons interferes with perceiving the visual signal clearly. The characteristic amount of this background variability added to the signal during processing is dubbed collectively as internal noise, which is known to provide an irreducible limit on detection. For example, the existence of an absolute contrast threshold that is higher than that of an ideal observer is evidence of such a limitation (Hecht et al., 1942; Rose, 1948; Barlow, 1956; Jones, 1959; Geisler, 2003).



Over the last several decades, visual psychophysicists have modeled the limitations inherent to the visual system in the contrast domain by measuring contrast thresholds for signals embedded in external noise (Burgess and Colborne, 1988; Pelli, 1990; Eckstein et al., 1997; Lu and Doshier, 1998; Solomon, 2002; Jeon et al., 2009; Klein and Levi, 2009). This method has been used to assay how the performance of visually normal adults is altered by changes in attention (Doshier and Lu, 2000; Lu et al., 2004) and by perceptual learning (Doshier and Lu, 1999; Gold et al., 1999; Doshier et al., 2004; Li et al., 2004). In addition, investigators have modeled changes occurring with development, both during childhood (Brown, 1994; Kiorpes and Movshon, 1998; Skoczenski and Norcia, 1998; Bogfjellmo et al., 2013; Falkenberg et al., 2014) and during aging (Pardhan et al., 1996; Betts et al., 2007), and the changes that occur after a history of early abnormal visual experience (Levi et al., 2007, 2008; Jeon et al., 2012; Falkenberg and Bex, 2014).

Typically, these approaches involve the manipulation of signals and experimentally controlled external noise in order to infer properties of the underlying perceptual process, which is presumably limited by the presence of various internal noise sources affecting perceptual sensitivity. By titrating the signal with the external noise, one can make inferences about how these internal noise sources affect sensory perception. When the relationship between the level of external noise and perceptual thresholds is measured and plotted in log-log coordinates, the resulting curve shows distinctive features, where the thresholds remain constant over the low external noise levels and then increase linearly as a function of external noise after a certain level of external noise. This curve is called the Threshold vs. Contrast or *TvC* curve. Systematic examination of the relative locations and shifts of *TvC* curves collected under different psychological manipulations or at different ages may reveal the underlying mechanisms responsible for changes in perceptual performance.

A few studies have used this approach to examine developmental changes in the levels of internal noise in the visual system of typically developing humans. Brown (1994) measured contrast detection and discrimination thresholds from infants whose age ranged from 49 to 51 days and adults using a minimum motion technique in which two gratings drifted in opposite directions. An observer used the participant's eye movements to determine the direction of the single grating (detection) or of the grating of higher contrast (discrimination). Threshold was defined as the minimum contrast difference for which the observer could make this determination accurately. The infants' thresholds were much more elevated for detection (factor of 50) than were their thresholds for discrimination (factor of 3). According to the modeling, the huge performance difference for detection between the infants and adults reflects higher intrinsic noise independent of stimulus contrast. Bogfjellmo et al. (2013) reached a similar conclusion when they measured sensitivity to the global direction of signal dots moving in a unitary direction against the noise dots moving in random directions.

Skoczenski and Norcia (1998) used visually evoked potentials (VEP) to record the electrophysiological responses of infants, aged 6–30 weeks, to sinusoidal gratings of varying contrast masked by varying amounts of temporally modulated noise. For

each external noise level chosen for testing, the contrast of the grating was diminished into the background gradually until no VEP response was elicited. This contrast level was considered to be the contrast threshold for that noise level and was plotted, along with the contrast thresholds measured at other noise levels, to create a *TvC* curve from which the authors estimated the infants' internal noise. They found that the amount of internal noise in newborns was approximately nine times that of adults tested in the same way. They observed a rapid decrease in internal noise between 6 and 10 weeks of age, after which time the infants' internal noise was only 1.8 times greater than that of adults even though the contrast thresholds of infants were still higher by a much greater factor. Since overall contrast thresholds improved over the same time period during which internal noise decreased, the authors suggest that internal noise is a major limitation on infants' contrast sensitivity.

The one developmental study using macaque monkeys also reported decreases in internal noise with age. Kiorpes and Movshon (1998) trained young monkeys aged 1–18 months and adult monkeys to pull a bar or look in the direction of a grating presented on the left or right side of a monitor. The grating was presented either with or without noise frames temporally alternating with the stimulus frames. The authors used the method of constant stimuli to determine a signal contrast threshold for each noise contrast, and each individual's amount of internal noise was estimated from the resulting *TvC* curve. In accordance with the findings of Skoczenski and Norcia (1998), Kiorpes and Movshon (1998) observed a decrease in both internal noise and contrast thresholds with age. However, the decrease in contrast threshold could not be explained completely by the changes in intrinsic noise.

Until recently (Falkenberg et al., 2014), the findings on the development of human contrast detection/discrimination in noise have been restricted to early infancy: the ages tested have ranged only from 6 to 30 weeks of age (Brown, 1994; Skoczenski and Norcia, 1998). Falkenberg and colleagues used an equivalent noise paradigm to investigate and model the development and maturation of motion perception (detection, summation, and discrimination) in school-aged children (5–14 years) and adults. Measuring contrast thresholds at only two levels of external noise (no noise and high noise), they found a long developmental trajectory for only the discrimination of the motion direction, for which the contrast thresholds decreased continually into the adolescence. The authors modeled the decrease as arising from an improvement in sampling efficiency with no change in internal noise.

The previous studies with humans compared changes in performance across age groups by measuring a single *TvC* curve for each age group obtained at only one performance criterion (e.g., 75% correct). Although comparing single *TvC* curves provides valuable information about internal noise, measuring only one *TvC* curve cannot capture fully the mechanisms underlying performance change. In their detailed explanations of this point, Lu and Doshier (1998, 1999, 2008) and Lu et al. (2004) argued and demonstrated in a series of papers that more than one *TvC* curve must be measured at different performance criteria in order to characterize satisfactorily the mechanisms underlying various

perceptual tasks. Measurement at multiple performance levels allows separate calculations of threshold ratios at each level of external noise. The additional data allow one to calculate separate estimates of internal additive noise, internal multiplicative noise, and template retuning (which is also called *excluding external noise*).

Measuring extra data points increases the power of the study, but also increases the time needed for testing, which can be especially problematic when testing children. To decrease the time required for data collection without sacrificing the quality of data, we used quick *TvC* (*qTvC*; Lesmes et al., 2006) to estimate contrast thresholds of four age groups (5-, 7-, 9-years-olds, and adults) at multiple performance levels across a wide range of embedded external noise levels. Specifically, we used *qTvC* to calculate contrast thresholds corresponding to 60, 75, and 90% correct performance levels for each of nine different external noise levels to obtain three *TvC* curves for each participant. With the aid of the *qTvC* method, we could collect data for each participant in less than 20 min. To model the mechanisms underlying developmental changes with age, we calculated an average *TvC* curve for each age group at each performance level. The current study is the first to evaluate the source of the known improvement in contrast thresholds with age in school-aged children (e.g., Ellemberg et al., 1999). To do so, we used the model of Lu and Doshier described above that has been successful in establishing the source of the limitations in adults tested at multiple levels of performance (Lu and Doshier, 1998, 1999, 2008; Lu et al., 2004).

## MATERIALS AND METHODS

### PARTICIPANTS

We tested four groups of participants: twenty 5.5-years-olds  $\pm$  3 months (mean age = 5.5 years,  $SD = 0.13$ , 5 female), twenty 7.5-year-olds  $\pm$  3 months (mean age = 7.5 years,  $SD = 0.11$ , 10 female), twenty 9.5-year-olds  $\pm$  3 months (mean age = 9.5 years,  $SD = 0.14$ , 11 female), and twenty adults ranging in age from 18.1 to 24.5 years (mean age = 19.6 years,  $SD = 1.45$ , 12 female). All participants in the final sample had passed a visual screening exam. Two additional 5.5-year-olds, one additional 7.5-year-old, and two additional adult participants were excluded because they did not pass the visual screening exam (see Section Procedure). The children were recruited from a database of children whose parents had volunteered to participate at the time of the child's birth. Children received a Junior Scientist certificate and a toy or a book voucher for their participation. Adults were volunteers or McMaster undergraduate psychology students who participated for course credit or \$10 compensation.

### APPARATUS AND STIMULI

The stimuli were presented on a 20 inch Sony Trinitron VGA color monitor with a pixel resolution of  $640 \times 480$  and a 100 Hz refresh rate. The test stimuli were created in MATLAB (Mathworks, 2008). A stimulus sequence in a trial lasted 90 ms and consisted of nine alternating 10 ms patches of signal and noise in the following sequence: noise1-signal1-noise2-...-noise4-signal4-noise5. The signal was a Gaussian-windowed sinusoidal Gabor with a spatial frequency of 1 c/deg oriented  $\pm 45^\circ$  from vertical. The alternation of the noise patches and the Gabor was fast enough that

the noise appeared to be superimposed spatially on the Gabor. The luminance profile of the Gabor stimulus is described by the following equation:

$$L(x, y) = L_0(1.0 + c \sin[2\pi f(x \cos \theta + y \sin \theta)]) \exp[-(x^2 + y^2)/2\sigma^2] \quad (1)$$

where  $c$  is signal contrast,  $\sigma$  is the standard deviation of the Gaussian window ( $1.86^\circ$ ),  $f$  is spatial frequency, and  $L_0$  is the background luminance which was set to the middle of the dynamic range of the display. The Gabor stimulus and noise patches were presented in a  $7.8^\circ \times 7.8^\circ$  frame when viewed from 57 cm. An external noise patch was composed of  $0.1^\circ \times 0.1^\circ$  pixel granules, the contrasts of which were sampled independently for each frame from a Gaussian distribution with a mean of 0 and one of the nine standard deviations (ranging from 2 to 33%, sampled in 3dB steps) as prescribed by the *qTvC*.

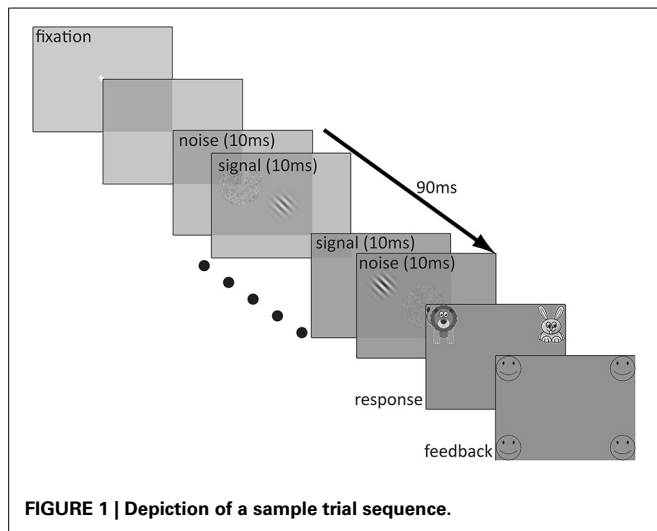
Before each trial, a white fixation cross ( $0.6^\circ \times 0.6^\circ$ , line width =  $0.062^\circ$ ) was presented in the center of the monitor for 500 ms, followed by 250 ms of blank screen prior to the onset of the test stimulus. Immediately after the test stimulus, there appeared a response screen which consisted of an image of a cartoon lion in the upper left corner of the screen (bottom edge  $3^\circ$  above center and inner edge  $8^\circ$  to the left of center) and an image of a cartoon rabbit in the upper right corner of the screen (bottom edge  $3^\circ$  above center and inner edge  $8^\circ$  to the right of center). The size of each image was  $12^\circ \times 12^\circ$  and a white question mark was centered between them. This screen remained indefinitely until the participant made a response. Participants indicated whether the top of the Gabor was tilted to the right (rabbit) or to the left (lion) by pressing a key (the F key on the left side of the keyboard to indicate that the stimulus was angled to the left and the J key on the right side of the keyboard to indicate the stimulus was angled to the right). If participants preferred, they responded verbally by saying "left" or "lion" for a leftward choice and "right" or "rabbit" for a rightward choice to a blind experimenter who entered the response on a keyboard. If participants chose the correct answer, they received positive feedback in the form of four outlined, circular smiley faces (each  $8^\circ$  in diameter), one in each of the four corners of the monitor (closest edges  $7^\circ$  above and below center and inner edges  $12^\circ$  to the left and right of center) and an encouraging cheering sound. However, if participants responded incorrectly, they received four outlined frowning faces of the same size and location and heard a "D'oh!" sound indicating that their choice was incorrect. A typical trial sequence is depicted in Figure 1.

### PROCEDURE

Prior to any procedures, we obtained informed consent from participants or their parents. We also obtained assent from the children 8 years and older. Adult participants and parents of children were provided with a debriefing form upon completion of the experiment. Our experimental procedures were cleared by the McMaster Research Ethics Board.

### Visual screening procedure

All participants in the final sample had normal or corrected-to-normal vision, for their age. The visual screening exam included



tests of linear letter acuity, binocular fusion, and stereo acuity. Adults, 9- and 7-year-olds were required to have a linear letter acuity enabling them to read correctly all but two letters on the 20/20 line in each eye when tested monocularly with the Lighthouse Distance Visual Acuity Test chart. The 5-year-olds had a linear letter acuity of at least 20/25 when tested with the Goodlite Crowding cards. If necessary, participants were given spectacle corrections of up to  $-1.5$  dioptres to ensure that any myopic error was too small to interfere with vision at the testing distance of 57 cm. Participants were required to have worse acuity with an added  $+3.00$  dioptre lens to rule out hypermetropia (farsightedness) of more than 3 dioptres. Binocular fusion was assessed using the Worth 4-Dot Test, and stereoacuity was assessed with the Titmus Fly Stereotest. Participants were required to show evidence of binocular fusion and stereoacuity of at least 100 arcsec for the 5-year-olds and 40 arcsec for the older participants.

### Experimental procedure

Participants sat in a darkened room and viewed the stimuli binocularly. The experimental procedure consisted of a demonstration, a criterion, a practice run, and a test run.

**Demonstration.** Participants were told that the pattern they were going to see “looks similar to a Ruffles® potato chip.” The experimenter showed participants a vertical Gabor with no noise and told them “the lines on the chip make it look like it is standing straight up and down.” This screen was presented until participants agreed verbally that the Gabor was oriented vertically. The experimenter then told participants that “the Ruffles® potato chip in the computer game will be tilted and the goal was to decide whether the top of the Ruffles® potato chip was tilted toward the right (toward the rabbit) or toward the left (toward the lion).” The experimenter then showed participants a static right-tilted Gabor in the center of a screen containing the previously described cartoon rabbit and lion and explained, “This is what the chip will look like when it is tilted toward the right (toward the rabbit).” They were shown this screen until they

agreed verbally that the Gabor was now tilted toward the rabbit. They were then shown the vertical Gabor again, followed by a static left-tilted Gabor on a screen containing the cartoon animals and were told that “This is what the chip will look like when it is tilted toward the left (toward the lion).” This screen was presented until participants agreed verbally that the Gabor was now tilted toward the lion. The participants were then asked to indicate verbally which way two practice Gabors were tilted. As in the first two trials, the two practice Gabors were presented without noise but were shown for only 1 s each. All participants responded correctly to these two 1-s practice trials and were given feedback by the computer program and by the experimenter.

**Criterion.** Next, we ensured that participants understood the task by testing them with a criterion session in which they were shown static Gabors at 50% contrast in no noise for 90 ms and were required to give four consecutive correct answers. The participants received feedback from the computer program for each of their responses. The experimenter then explained that sometimes the “Ruffles® potato chip will be sprinkled with salt and pepper and will look fuzzy.” On the computer screen, we presented one right-tilted Gabor alternated with a 50% contrast noise patch and then a similar left-tilted Gabor with 50% noise and told participants which way they were tilted. Both Gabors were presented until the participant indicated that they saw which way the “Ruffles® potato chip” was tilted beneath the salt and pepper. They were then tested with a second criterion session in which the Gabors were presented for 90 ms with 50% noise. Again, they were required to give four consecutive correct answers. Participants were required to pass each of the criterion sessions in no more than three blocks of four trials, and all did so in the first or second block.

**Practice.** After participants passed both criterion sessions successfully, the experimenter presented a 24-trial practice run. The stimuli used in the 24-trial practice run were generated by the *qTVC* program and were identical to the first 24 trials used in the test run. The computer program generated the three parameters of the resulting *TvC* curve based on the 24 practice trials: the critical noise ( $N_c$ ), the optimal contrast threshold ( $C_0$ ), and the common slope of the psychometric function ( $\eta$ ) and these parameters were recorded by the experimenter.

**Test run.** The test run was identical to the practice run except that it consisted of 240 trials differing in stimulus contrast and noise levels, as generated by the *qTVC* paradigm. The stimulus space for the *qTVC* procedure included nine possible external noise contrasts ranging from 2 to 33% (in 3dB steps), and signal contrast levels sampled from a pool of 40 possible contrast levels ranging from 0 to 90% (in 1dB steps). Participants who requested a break were given a 5-min quiet break in the testing room. At the end of the test run, the program reported thresholds corresponding to three levels of accuracy: 60, 75, and 90%. Each experimental session lasted approximately 40 min plus approximately 5 min for visual screening.

## MODELING

To quantify and model the improvement with age, we adopted the original Perceptual Template Model (PTM; Lu and Doshier, 1998), developed previously to characterize changes in perceptual performance with attention (Lu and Doshier, 1998; Doshier and Lu, 2000) and perceptual learning (Doshier and Lu, 1999; Lu et al., 2004). A detailed description of the PTM can be found in one of the cited papers. Briefly, overall performance of an observer, expressed in  $d'$ , is limited by the following three noise sources in the PTM: (1) external noise ( $N_{ext}$ ), the strength of which is known to the experimenter, (2) internal additive noise ( $N_{add}$ ), an irreducible amount of variability inside a system determining the lower bound on performance (Barlow, 1956; Pelli, 1990), and (3) multiplicative noise ( $N_{mul}$ ), an independent noise source, the strength of which is proportional to the stimulus strength (Green and Swets, 1974; Legge and Foley, 1980). The initial signal and noise composite may be subject to a non-linearity ( $\gamma$ ) in the system (Nachmias and Sansbury, 1974; Kontsevich et al., 2002). Combined, the overall performance of a system is fundamentally determined by the signal-to-noise ratio:

$$d' = \frac{S}{N_{total \text{ noise sources}}} = \frac{(\beta c)^\gamma}{\sqrt{N_{ext}^{2\gamma} + N_{mul}^2[(\beta c)^{2\gamma} + N_{ext}^{2\gamma}] + N_{add}^2}} \quad (2)$$

where  $c$  is the contrast of the signal and  $\beta$  represents the gain or amplification factor on the signal after a perceptual template which is tuned to the relevant dimension of stimulation (e.g., contrast in the current case). Rearranging the equation for the contrast threshold yields,

$$c_\tau = \frac{1}{\beta} \left[ \frac{(1 + N_{mul}^2) N_{ext}^{2\gamma} + N_{add}^2}{\left(\frac{1}{d'^2} - N_{mul}^2\right)} \right]^{\frac{1}{2\gamma}} \quad (3)$$

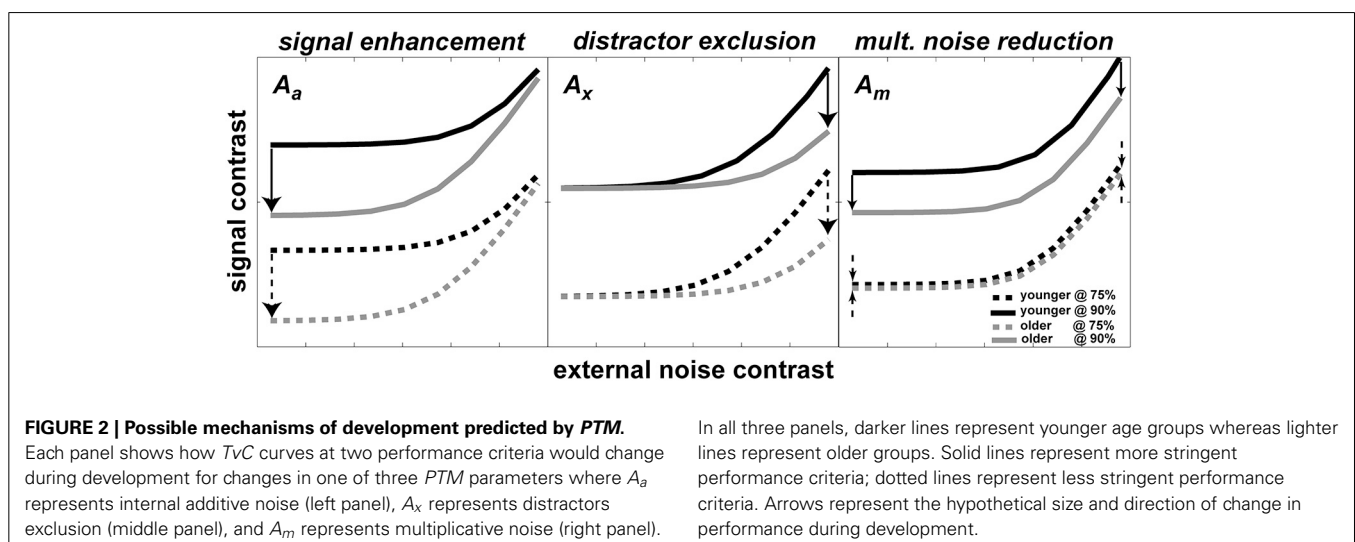
According to the PTM, improvement in performance through development can be modeled by changes in one or more of the noise sources. Each panel in **Figure 2** shows a hypothetical pattern of performance change when only one of the three mechanisms mentioned above is in operation: (1) *stimulus enhancement* (left panel)—represents improvement caused by a reduction in internal additive noise. In this case, the improvement will be shown in the low external noise region. (2) *External noise exclusion* (middle panel)—represents the ability to suppress or filter out irrelevant information (i.e., external noise). As opposed to case (1), this pattern of improvement will be shown when external noise is high. (3) *Internal multiplicative noise reduction* (right panel)—a reduction of internal multiplicative noise will improve performance over the entire range of external noise levels.

Assuming all three mechanisms are at work, we can rewrite the above equation to accommodate the developmental changes in our current data as following:

$$c_\tau = \frac{1}{\beta} \left[ \frac{\left(1 + (A_m^{(i)} N_{mul})^2\right) (A_x^{(i)} N_{ext})^{2\gamma} + (A_a^{(i)} N_{add})^2}{\left(\frac{1}{d'^2} - (A_m^{(i)} N_{mul})^2\right)} \right]^{\frac{1}{2\gamma}} \quad (4)$$

where the index ( $i$ ) denotes the age group. To quantify the relative contributions from each or combinations of the noise sources to the improvements with age, three extra coefficients  $A$ s with subscripts corresponding to each noise source are used. In this form, the relative improvements with age are quantified against the performance of 5-year-olds, where we set  $A_a^{5yro} = A_x^{5yro} = A_m^{5yro} = 1$ .

**Figure 2** also illustrates the important property of the PTM for contrast thresholds predicted at two different performance criteria for each mechanism. In each panel, there are two pairs of darker and lighter lines representing hypothetical  $TvC$  curves for younger and older observers, respectively. Each pair of curves was drawn at two different performance criteria (e.g., solid lines represent 90% correct performance level and broken lines represent 75% correct performance level). The direction of the arrows





represents the direction of improvement and the size of the arrows approximately matches the magnitude of improvement. Inspection of the figure highlights how the magnitude of change can be contingent upon performance criteria. In the cases of signal enhancement (left panel) and distractor exclusion (middle panel), for example, the size of improvement is constant regardless of the criteria. On the other hand, the size of improvement increases as the performance criterion becomes more stringent in the case of multiplicative noise reduction (right panel). Therefore, measuring multiple *TvC* provides strong constraints and is useful in distinguishing between mixtures of mechanism in the hierarchical model testing in the *PTM*.

## RESULTS

**Figure 3** shows the *TvC* curves at 75% correct performance for each age group (5-, 7-, 9-year-olds, and adults in red, green, blue, and black respectively). The left panel shows the individual outputs ( $n = 20$ /group) after running 240 trials of *qTvC*. The right panel shows the mean for each age group with the shaded regions representing  $\pm 1$  s.e.m. at each noise level. As age increases, the average performance improves (shown as decreases in contrast thresholds) over the entire noise range tested. The improvement seems greatest, especially in the high noise region, between 5 and 7 years of age, after which the improvement with age becomes more gradual. This pattern is evident, even when one takes into account the greater variability in performance at age 5.

**Figure 4** shows the developmental data over three different performance levels. As mentioned in the Introduction and the Section Modeling above, multiple *TvCs* at different criteria provides stronger constraints in distinguishing the mixtures of mechanisms (Doshier and Lu, 1999; Klein and Levi, 2009). Qualitatively, there is an increase in the threshold ratios among different age groups across the entire external noise range as a more stringent performance criterion is implemented, implying the impact of multiplicative noise on the change in contrast thresholds. From the disproportionate changes in threshold ratios

between the low and high external noise region across different performance criteria, we can also infer that signal enhancement and external noise exclusion may be at work at the same time.

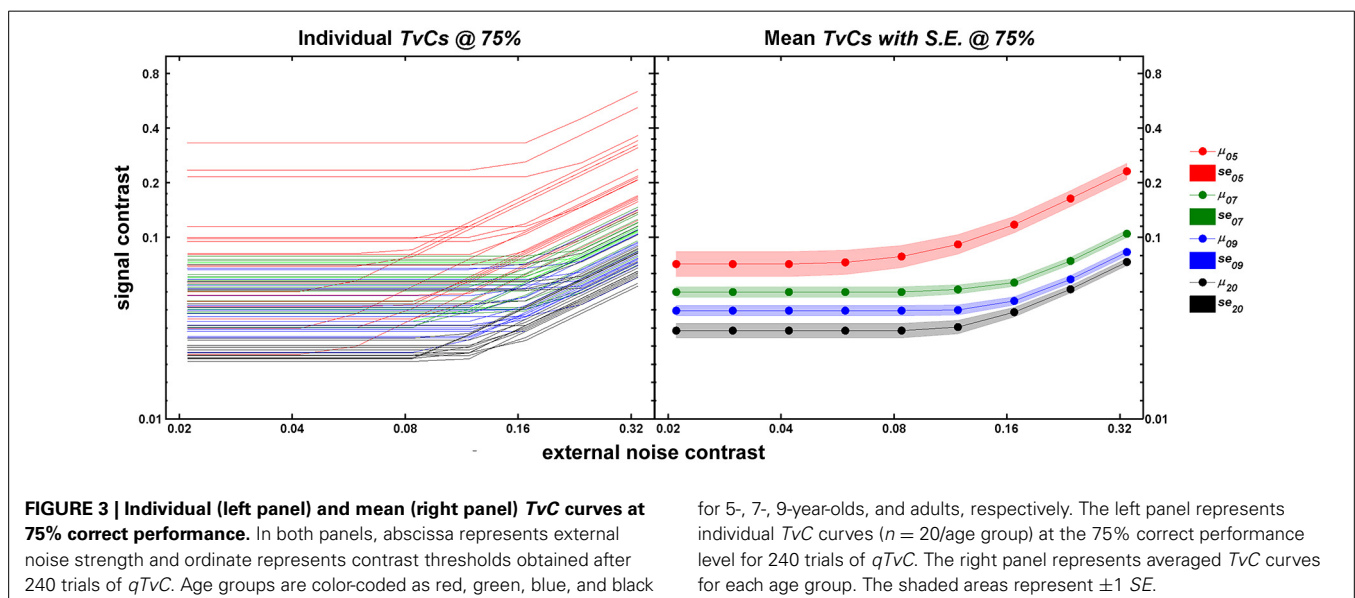
**Figure 5** shows the mean data for each age group and the result of nested model fitting of these mean data using the equation (4). In this figure, age group is arranged column-wise while different models used for fitting are arranged in rows. Each panel contains data (shown as dots) at three performance levels (60, 75, and 90% correct) with error bars and the resulting model fits shown as lines. The age-related improvements can be seen across the columns as a gradual decrease in thresholds regardless of the performance levels. Note that the distance between the contrast thresholds at different performance levels is distinctively wider in 5-year-olds' data than the data from the remaining age groups.

The total number of data points used in this fitting procedure was 108 (9 noise levels  $\times$  3 performance criteria  $\times$  4 age groups). There are four layers of models with each layer of the same model containing the same number of parameters. For example, the most saturated layer has a model with 13 parameters (denoted as “full” in **Figure 5**) whereas the most parsimonious layer has a model with only four parameters (“no change” in **Figure 5**). There are a total of eight possible models across layers. With each model, we calculated goodness-of-fit ( $r^2$ ) (Equation 5) and compared them statistically (Equation 6) between layers.

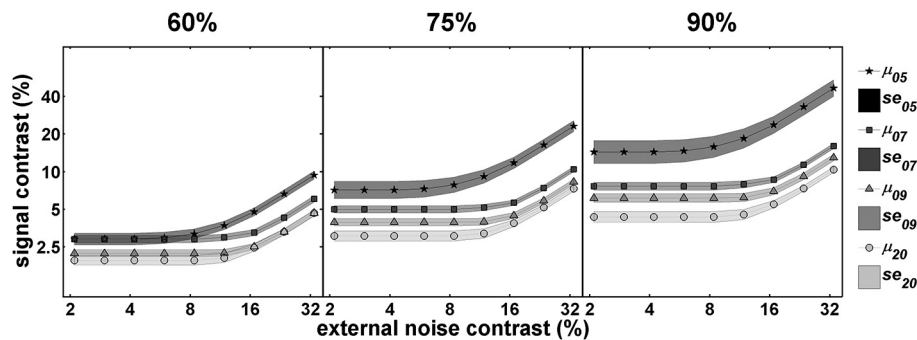
$$r^2 = 1 - \frac{\sum [\log(c_{\tau}^{PTM}) - \log(c_{\tau}^{data})]^2}{\sum [\log(c_{\tau}^{data}) - \text{mean}(\log(c_{\tau}^{data}))]^2} \quad (5)$$

Of all 22 comparisons, no models in the sub-layers produced statistically equivalent goodness-of-fit compared to the goodness-of-fit for the most saturated model in the top layer with 13 parameters (top row in **Figure 5**, highlighted with boldface).

$$F(df_1, df_2) = \frac{(r_{upper}^2 - r_{lower}^2)/df_1}{(1 - r_{upper}^2)/df_2} \quad (6)$$







**FIGURE 4 | Mean TvC curves at three performance criteria.** This figure illustrates how the magnitude of the change in performance with age, depicted by relative shifts among TvC curves, can be dependent upon performance criteria. Illustrations are for performance

criteria of 60% (left panel), 75% (middle panel), and 90% (right panel). In all panels darker lines represent younger age groups whereas lighter lines represent older groups. See text for more details.

where  $df_1 = k_{upper} - k_{lower}$ , and  $df_2 = N - k_{upper}$ . The  $k$ s are the number of parameters in each model to fit the data, and  $N$  is the number of data points to fit. We calculated bootstrapped confidence intervals for the best fitting PTM parameters by fitting the model 1000 times to the synthetic TvC thresholds resampled from each of three  $qTvC$  parameter distributions obtained from our observers. The pair of number in parentheses represents the confidence interval for each parameter. The full list of model parameters can be found in Appendix A in Supplementary material and the complete results of nested model comparisons are provided in Appendix B in Supplementary material.

The best model (top row in Figure 5) provided an excellent fit ( $r^2 = 0.985$ ) to the contrast thresholds at multiple levels of performance (60, 75, and 90%) across a wide range of external noise levels. The model suggests that a mixture of mechanisms underlies the developmental changes: the improvements in contrast thresholds over ages were best modeled by a combination of reductions in internal additive and multiplicative noise and improvements in excluding external noise (see Table 1). In line with the data, the improvement was greatest between 5 and 7 years of age, accompanied by a 38.6% reduction in additive noise, 70.7% reduction in multiplicative noise, and 45.1% improvement in external noise exclusion.

Figure 6 shows relative changes in each noise source with age. While both internal additive noise ( $A_a$ ) and the ability to exclude distractors ( $A_x$ ) seem to reach adult levels at the age of 9, multiplicative noise continues ( $A_m$ ) to decrease after age 9 (the oldest age of child tested here).

## DISCUSSION

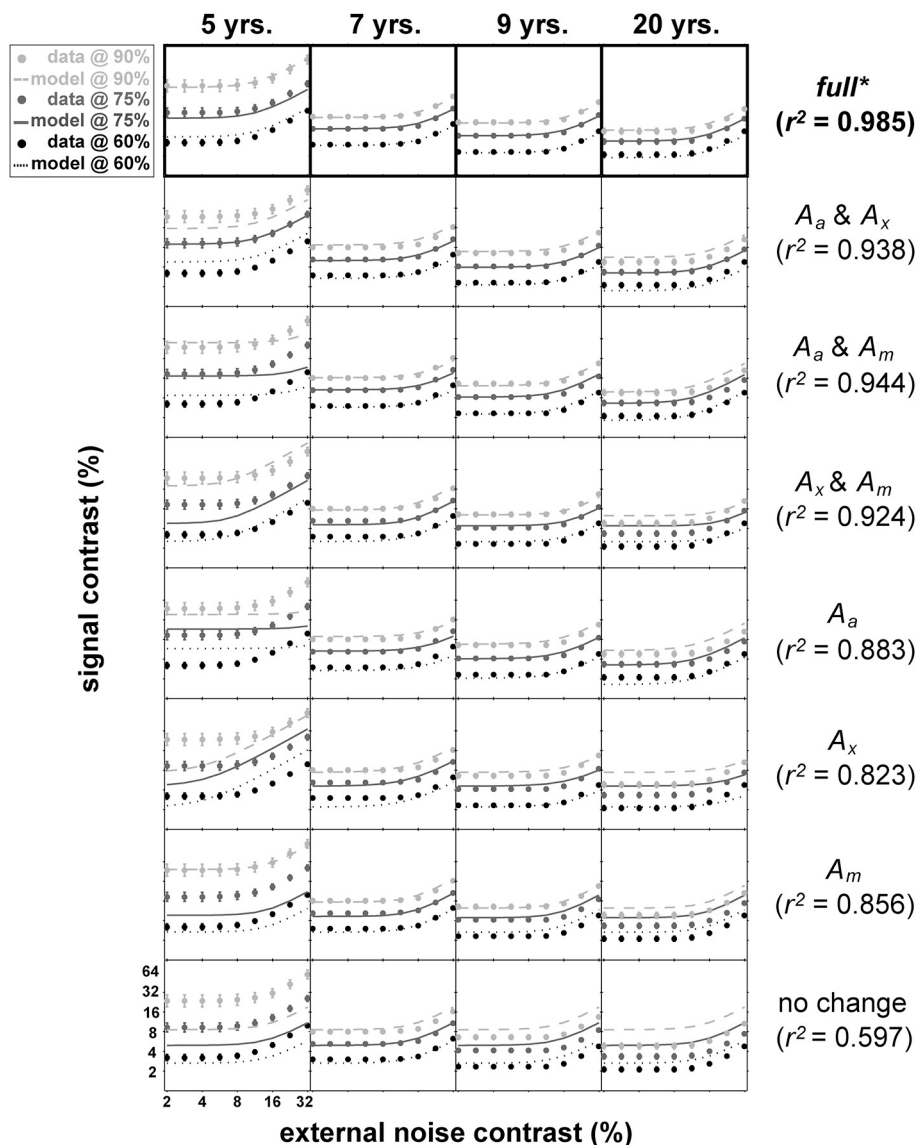
The purpose of the current study was to measure contrast thresholds embedded in a wide range of external noise in four age groups and to model the developmental improvements in contrast thresholds in terms of changes in limiting factors affecting visual performance. With a  $qTvC$  procedure (Lesmes et al., 2006), contrast thresholds at multiple performance criteria across nine external noise levels were estimated quickly in children and adults. We modeled our data with PTM to investigate whether the developmental improvement in contrast threshold with age can be

modeled by a combination of reduction in internal additive and multiplicative noise components as well as the improvement in filtering out irrelevant information.

In a previous study (Jeon et al., 2012), we included a task similar to the current experiment as one of the outcome measures to gauge the effect of video game training on the vision of adult congenital cataract patients and normal adult controls. In doing so, we applied the  $qTvC$  for the first time to collect 240 trials of data before and after the video game training. In the current study, we were able to collect data on 80 observers from a broad age range, highlighting the efficiency of  $qTvC$  in measuring and specifying the performance space defined across a wide range of noise and signal intensity.

Previous developmental studies consistently reported that infants and children are worse than adults at detecting or discriminating signals embedded in noise (Brown, 1994; Skoczenski and Norcia, 1998; Falkenberg et al., 2014). Those studies with infants found that the immaturity could be explained by higher internal additive noise. On the other hand, Falkenberg et al. (2014) found that poor sampling efficiency is responsible for the immaturity in motion discrimination of children and adolescents while the internal noise played no role in the development of motion discrimination after age 5, the youngest age tested. Bogfjellmo et al. (2013) reached a similar conclusion about sensitivity to the global direction of signal dots moving in a unitary direction against noise dots moving in random directions. Our work contrasts with these previous studies because we used a method that allowed us to distinguish between additive and multiplicative internal noise. At least for our task (contrast thresholds for orientation discrimination), internal additive noise was higher than in adults as late as age 7 and internal multiplicative noise was higher even at age 9, the oldest group of children tested. Specifically, the model identified three limits on 5-year-olds' contrast thresholds: (1) internal additive noise, (2) internal multiplicative noise, and (3) insufficient filtering of external noise.

First, our model showed that internal additive noise decreases with age for measurements of orientation discrimination in the contrast domain. Compared to 5-year-olds, there is a 39% reduction in 7-year-olds, a 60% reduction in 9-year-olds, and a 70%



**FIGURE 5 | Results of nested modeling.** This figure matrix shows the results of nested *PTM* modeling arranged by age (columns) and model layers (rows). In each panel, dots represent mean contrast thresholds at each noise level tested using *qTVC*, and smooth curves represent model fit to the data. The performance criteria become less stringent

as the colors become darker. Next to the figure matrix, goodness-of-fit statistics for a given model layer are provided. Statistical comparisons among model layers reveal that the “full” model in the top row is the best model, highlighted by an asterisk and thick black outlines around the top panels.

reduction in adults. These reductions account for the improvements in performance in the low external noise region.

Second, our data showed that the age-related change in contrast thresholds is dependent upon performance criteria, which is indicative of change in the level of internal multiplicative noise. As illustrated in **Figure 4**, the performance difference among age groups increased when higher accuracy was required. According to our modeling results, internal multiplicative noise also decreases with age. Compared to 5-year-olds, there is a 48–71% reduction by age 7–9 of internal multiplicative noise, and a complete elimination of it in adults, corresponding to a reduction of nearly 100%. There are competing points of view on

what is responsible for the rising thresholds with increasing noise, masking, or pedestal values: multiplicative noise vs. contrast gain control. Empirically the influence of multiplicative noise is indistinguishable from that of a contrast-gain control mechanism (Dao et al., 2006; Klein and Levi, 2009; Chen et al., 2014). In a developmental study of contrast gain control using VEP (Garcia-Quispe et al., 2009), human infants from 15 to 28 weeks showed little contrast gain control compared to the older observers. This is the first study to make measurements of this factor in older children. The continuous reduction of multiplicative noise throughout childhood shown in our study might suggest a long developmental trajectory in the contrast gain control mechanism.

**Table 1 | PTM parameter outputs from the best model.**

Parameter \ Age	5	7	9	Adult
$\beta$		2.1892 (1.8276, 2.5507)*		
$\gamma$		1.6778 (1.5144, 1.8412)		
$N_{add}$		0.0315 (0.0215, 0.0416)		
$N_{mul}$		0.3763 (0.3459, 0.4067)		
$A_a$	1	0.6139 (0.33, 0.8979)	0.3966 (0.1418, 0.6511)	0.3001 (0.1224, 0.4778)
$A_x$	1	0.5492 (0.4295, 0.6690)	0.4229 (0.3308, 0.5150)	0.3998 (0.3250, 0.4745)
$A_m$	1	0.2939 (-0.2066, 0.7934)	0.5161 (0.0163, 1.0163)	0.0004 (0.0004, 0.0005)

$\beta$ , gain from template matching.

$\gamma$ , non-linearity exponent.

$N_{add}$ , standard deviation of additive noise.

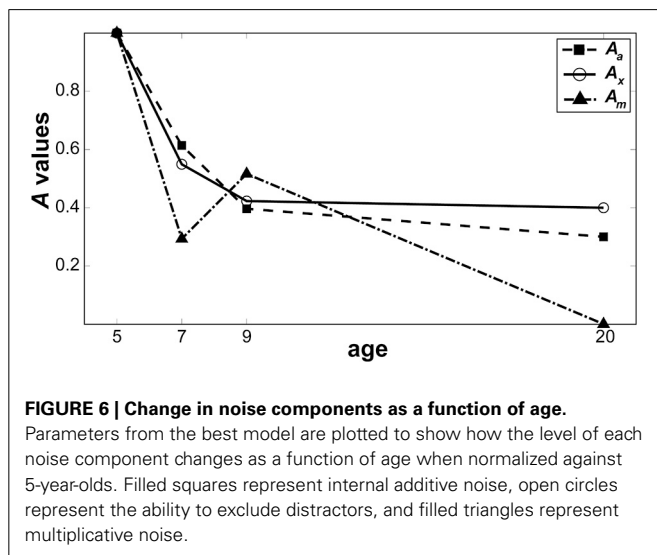
$N_{mul}$ , standard deviation of multiplicative noise.

$A_a$ , a developmental parameter associated with signal enhancement.

$A_x$ , a developmental parameter associated with external noise exclusion.

$A_m$ , a developmental parameter associated with internal multiplicative noise reduction.

\*The numbers in parentheses represent bootstrapped confidence intervals for the best PTM parameters by fitting the synthetic thresholds resampled 1000 times from the three qTVC parameter distributions obtained from our observers.



Alternatively, or in addition, it might reflect a long developmental trajectory for the reduction in multiplicative noise.

A third factor responsible for the age-related improvements we observed in contrast threshold was an improvement in the ability to filter out external noise, which is reflected as improvements in contrast thresholds at high external noise levels. Compared to 5-year-olds, the impact of the external noise on discrimination was reduced by 45% in 7-year-olds, 58% in 9-year-olds, and 60% in adults. Studies of perceptual learning (Lu and Doshier, 1999; Chung et al., 2005), and aging (Betts et al., 2007) confirm that performance can be improved by increased exclusion of external noise, achieved by retuning an internal template to the stimulus

property relevant to a given task so that it filters out incoming noise.

During development, channel reweighting (e.g., Lu and Doshier, 1999) of the sensory inputs likely becomes increasingly selective and tuned to the most relevant channel for forming perceptual decisions for a given task. Thus, given the shallow slope of the psychometric function in 5-year-olds, their response might be more similar than that at older ages across a wider range of input signals, the strength of which varies with external noise. This, in turn, would lower the differential signal-to-noise ratios around the relevant channels. For the visual system of 5-year-olds, this insensitivity to contrast might make it difficult to choose selectively the optimal channel for discrimination. In fact, substantial evidence indicates that young children are not optimal in selecting and processing the visual information that is most relevant to a given task. For example, the literature on visual selective attention indicates that children are not as good as adults at filtering out irrelevant background stimuli (Enns and Girgus, 1985; Ridderinkhof and Van Der Molen, 1995; Goldberg et al., 2001), with children as old as 10 years being affected more by distractors than adults (Goldberg et al., 2001). As reported by our best model output (Figure 6), it seems that the ability to cull external noise improves continually until 9 years of age.

Even though physiological changes such as pruning of excessive synaptic connections within the primary visual cortex, still occur until early adolescence (Huttenlocher et al., 1982; Garey and De Courten, 1983), it is unclear how much front-end changes in the structure or morphology of the early visual pathway can account for the developmental changes observed in our current age groups. In their study evaluating the developmental changes in contrast threshold and intrinsic noise using infant monkeys, Kiorpes and Movshon (1998) argued that changes in both additive and non-additive sources of noise contribute to the fall of the contrast thresholds during development. To arrive at this conclusion, they considered additive noise to represent the limiting factors in the early visual pathways and non-additive noise to represent “central” limiting factors, which might be tantamount to our internal multiplicative noise reduction and distractor exclusion. The documented changes in the striate visual pathway that continue well into adolescence may be responsible for such changes (Shaw et al., 2008; Pinto et al., 2010).

Even though the length of our procedure was reduced with the aid of qTVC, it might still be possible that children are simply less motivated or have a poorer understanding of the task. However, it is unlikely that worse performance in younger age groups was caused by a lack of motivation or understanding. First, we made sure that children understood the task by showing them demonstration trials, documenting their understanding with criterion trials, and familiarizing them with the test by having them complete a full session of qTVC before the data to be used were collected. Second, we kept the children motivated throughout the task by adding humorous auditory feedback when the child answered correctly. Although they were told that they could stop at any time, no child decided to discontinue the study, and all children seemed to enjoy the experimental procedure. Third, the qTVC algorithm kept performance much higher than chance level on most trials. Therefore, our observed effects were most likely a

consequence of factors related to visual sensitivity and minimally affected by cognitive immaturity or lack of motivation.

In summary, the results from the current study suggest that the contrast sensitivity of 5-year-olds is limited by higher levels of internal additive and multiplicative noise and higher susceptibility to irrelevant background information. There are rapid decreases in these limitations until age 7 and gradual reductions thereafter, with the reduction in multiplicative noise continuing past age 9, the oldest age tested here. It can be hypothesized that these limitations at age 5 can explain previous observations of poorer thresholds and decreased psychometric slopes compared to older ages. Our model using a mixture of reductions in internal additive noise, reductions in internal multiplicative noise, and an improvement in the ability to filter out external noise can account well for the age-related improvements in contrast threshold.

## ACKNOWLEDGMENTS

Parts of the data were presented at the annual meetings of the Society for Neuroscience, San Diego, CA, November 2010 and the Vision Sciences Society, Naples, FL, May 2012. We thank Jennifer Weeks for her help in collecting and analyzing the data and Sally Stafford for her help in booking and testing some of the participants. This research was supported by grants to authors Daphne Maurer and Terri L. Lewis from the Canadian Institutes of Health Research (MOP 36430).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00977/abstract>

## REFERENCES

- Banks, M. S., and Bennett, P. J. (1988). Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. *J. Opt. Soc. Am. A* 5, 2059–2079. doi: 10.1364/JOSAA.5.002059
- Barlow, H. B. (1956). Retinal noise and absolute threshold. *J. Opt. Soc. Am. A* 46, 634–639.
- Betts, L. R., Sekuler, A. B., and Bennett, P. J. (2007). The effects of aging on orientation discrimination. *Vision Res.* 47, 1769–1780. doi: 10.1016/j.visres.2007.02.016
- Bogfjellmo, L.-G., Bex, P. J., and Falkenberg, H. K. (2013). Reduction in direction discrimination with age and slow speed is due to both increased internal noise and reduced sampling efficiency. *Invest. Ophthalmol. Vis. Sci.* 54, 5204–5210. doi: 10.1167/iops.13-12005
- Braddick, O., and Atkinson, J. (2011). Development of human visual functions. *Vision Res.* 51, 1588–1609. doi: 10.1016/j.visres.2011.02.018
- Brown, A. M. (1994). Intrinsic contrast noise and infant visual contrast discrimination. *Vision Res.* 34, 1947–1964. doi: 10.1016/0042-6989(94)90025-6
- Burgess, A. E., and Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *J. Opt. Soc. Am. A* 5, 617.
- Chandna, A. (1991). Natural history of the development of visual acuity in infants. *Eye (Lond)*. 5(Pt 1), 20–26. doi: 10.1038/eye.1991.4
- Chen, G., Hou, F., Yan, F.-F., Zhang, P., Xi, J., Zhou, Y., et al. (2014). Noise provides new insights on contrast sensitivity function. *PLoS ONE* 9:e90579. doi: 10.1371/journal.pone.0090579
- Chung, S. T., Levi, D. M., and Tjan, B. S. (2005). Learning letter identification in peripheral vision. *Vision Res.* 45, 1399–1412. doi: 10.1016/j.visres.2004.11.021
- Cohn, T. E., and Lasley, D. J. (1986). Visual sensitivity. *Annu. Rev. Psychol.* 37, 495–521. doi: 10.1146/annurev.ps.37.020186.002431
- Dao, D., Lu, Z.-L., and Doshier, B. A. (2006). Adaptation to sine-wave gratings selectively reduces the contrast gain of the adapted stimuli. *J. Vis.* 6, 739–759. doi: 10.1167/6.7.6
- Doshier, B. A., Liu, S.-H., Blair, N., and Lu, Z.-L. (2004). The spatial window of the perceptual template and endogenous attention. *Vision Res.* 44, 1257–1271. doi: 10.1016/j.visres.2004.01.011
- Doshier, B. A., and Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Res.* 39, 3197–3221. doi: 10.1016/S0042-6989(99)00059-0
- Doshier, B. A., and Lu, Z. L. (2000). Mechanisms of perceptual attention in precueing of location. *Vision Res.* 40, 1269–1292. doi: 10.1016/S0042-6989(00)00019-5
- Eckstein, M. P., Ahumada, A. J., and Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *J. Opt. Soc. Am. A* 14, 2406–2419.
- Ellemberg, D., Lewis, T. L., Liu, C. H., and Maurer, D. (1999). Development of spatial and temporal vision during childhood. *Vision Res.* 39, 2325–2333. doi: 10.1016/S0042-6989(98)00280-6
- Enns, J. T., and Gireus, J. S. (1985). Developmental changes in selective and integrative visual attention. *J. Exp. Child Psychol.* 40, 319–337. doi: 10.1016/0022-0965(85)90093-1
- Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). Noise in the nervous system. *Nat. Rev. Neurosci.* 9, 292–303. doi: 10.1038/nrn2258
- Falkenberg, H. K., and Bex, P. J. (2014). Sources of motion-sensitivity loss in glaucoma. *Invest. Ophthalmol. Vis. Sci.* 48, 2913–2921. doi: 10.1167/iops.06-0752
- Falkenberg, H. K., Simpson, W. A., and Dutton, G. N. (2014). Development of sampling efficiency and internal noise in motion detection and discrimination in school-aged children. *Vision Res.* 100, 8–17. doi: 10.1016/j.visres.2014.04.001
- Garcia-Quispe, L., Gordon, J., and Zemon, V. (2009). Development of contrast mechanisms in humans: a VEP study. *Optom. Vis. Sci.* 86, 708–716. doi: 10.1097/OPX.0b013e3181a61673
- Garey, L. J., and De Courten, C. (1983). Structural development of the lateral geniculate nucleus and visual cortex in monkey and man. *Behav. Brain Res.* 10, 3–13. doi: 10.1016/0166-4328(83)90145-6
- Geisler, W. S. (2003). “Ideal observer analysis,” in *Visual Neurosciences*, eds L. M. Chalupa and J. S. Werner (Boston, MA: MIT Press), 825–837.
- Gold, J., Bennett, P. J., and Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature* 402, 176–178. doi: 10.1038/46027
- Goldberg, M., Maurer, D., and Lewis, T. L. (2001). Developmental changes in attention: the effects of endogenous cueing and of distractors. *Dev. Sci.* 4, 209–219. doi: 10.1111/1467-7687.00166
- Green, D. M., and Swets, J. A. (1974). *Signal Detection Theory*. New York, NY: Wiley.
- Hecht, S., Shlaer, S., and Pirenne, M. H. (1942). Energy, quanta, and vision. *J. Gen. Physiol.* 25, 819–840. doi: 10.1085/jgp.25.6.819
- Huttenlocher, P. R., De Courten, C., Garey, L. J., and Van Der Loos, H. (1982). Synaptogenesis in human visual cortex – evidence for synapse elimination during normal development. *Neurosci. Lett.* 33, 247–252. doi: 10.1016/0304-3940(82)90379-2
- Jeon, S. T., Lu, Z. L., and Doshier, B. A. (2009). Characterizing perceptual performance at multiple discrimination precisions in external noise. *J. Opt. Soc. Am. A* 26, B43–B58. doi: 10.1364/JOSAA.26.000B43
- Jeon, S. T., Maurer, D., and Lewis, T. L. (2012). The effect of video game training on the vision of adults with bilateral deprivation amblyopia. *Seeing Perceiving* 25, 493–520. doi: 10.1163/18784763-00002391
- Jones, R. C. (1959). Quantum efficiency of human vision. *J. Opt. Soc. Am. A* 49, 645–653.
- Kiorpes, L., and Movshon, J. A. (1998). Peripheral and central factors limiting the development of contrast sensitivity in Macaque monkeys. *Vision Res.* 38, 61–70. doi: 10.1016/S0042-6989(97)00155-7
- Kiorpes, L., Tang, C., Hawken, M. J., and Movshon, J. A. (2003). Ideal observer analysis of the development of spatial contrast sensitivity in macaque monkeys. *J. Vis.* 3, 630–641. doi: 10.1167/3.10.6
- Klein, S. A., and Levi, D. M. (2009). Stochastic model for detection of signals in noise. *J. Opt. Soc. Am. A* 26, B110–B126. doi: 10.1364/JOSAA.26.000B110
- Kontsevich, L. L., Chen, C. C., and Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Res.* 42, 1771–1784. doi: 10.1016/S0042-6989(02)00091-3
- Legge, G. E., and Foley, J. M. (1980). Contrast masking in human vision. *J. Opt. Soc. Am. A* 70, 1458–1471.
- Lesmes, L. A., Jeon, S.-T., Lu, Z.-L., and Doshier, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: the quick TvC method. *Vision Res.* 46, 3160–3176. doi: 10.1016/j.visres.2006.04.022

- Levi, D. M., Klein, S. A., and Chen, I. (2007). The response of the amblyopic visual system to noise. *Vision Res.* 47, 2531–2542. doi: 10.1016/j.visres.2007.06.014
- Levi, D. M., Klein, S. A., and Chen, I. (2008). What limits performance in the amblyopic visual system: seeing signals in noise with an amblyopic brain. *J. Vis.* 8, 1–23. doi: 10.1167/8.4.1
- Li, R. W., Levi, D. M., and Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision ‘template’ for position discrimination. *Nat. Neurosci.* 7, 178–183. doi: 10.1038/nn1183
- Lu, Z. L., and Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Res.* 38, 1183–1198. doi: 10.1016/S0042-6989(97)00273-3
- Lu, Z. L., and Doshier, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *J. Opt. Soc. Am. A* 16, 764–778. doi: 10.1364/JOSAA.16.000764
- Lu, Z. L., and Doshier, B. A. (2008). Characterizing observers using external noise and observer models: assessing internal representations with external noise. *Psychol. Rev.* 115, 44–82. doi: 10.1037/0033-295X.115.1.44
- Lu, Z. L., Jeon, S. T., and Doshier, B. A. (2004). Temporal tuning characteristics of the perceptual template and endogenous cuing of spatial attention. *Vision Res.* 44, 1333–1350. doi: 10.1016/j.visres.2003.12.017
- Mathworks (2008). *MATLAB 7.6*. Natick, MA: The Mathworks Inc.
- Nachmias, J., and Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Res.* 14, 1039–1042. doi: 10.1016/0042-6989(74)90175-8
- Neu, B., and Sireteanu, R. (1997). Monocular acuity in preschool children: assessment with the Teller and Keeler acuity cards in comparison to the C-test. *Strabismus* 5, 185–202. doi: 10.3109/09273979709044534
- Norcia, A. M., and Tyler, C. W. (1985). Infant VEP acuity measurements: analysis of individual differences and measurement error. *Electroencephalogr. Clin. Neurophysiol.* 61, 359–369. doi: 10.1016/0013-4694(85)91026-0
- Pardhan, S., Gilchrist, J., Elliott, D. B., and Beh, G. K. (1996). A comparison of sampling efficiency and internal noise level in young and old subjects. *Vision Res.* 36, 1641–1648. doi: 10.1016/0042-6989(95)00214-6
- Pelli, D. G. (1990). “The quantum efficiency of vision,” in *Vision: Coding and Efficiency*, ed C. Blakemore (Cambridge: Cambridge University Press), 3–24.
- Pinto, J. G. A., Hornby, K. R., Jones, D. G., and Murphy, K. M. (2010). Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan. *Front. Cell. Neurosci.* 4:16. doi: 10.3389/fncel.2010.00016
- Ridderinkhof, K. R., and Van Der Molen, M. W. (1995). When global information and local information collide: a brain potential analysis of the locus of interference effects. *Biol. Psychol.* 41, 29–53. doi: 10.1016/0301-0511(95)05125-T
- Rose, A. (1948). The sensitivity performance of the human eye on an absolute scale. *J. Opt. Soc. Am. A* 38, 196–208.
- Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *J. Neurosci.* 28, 3586–3594. doi: 10.1523/JNEUROSCI.5309-07.2008
- Skoczenski, A. M., and Norcia, A. M. (1998). Neural noise limitations on infant visual sensitivity. *Nature* 391, 697–700. doi: 10.1038/35630
- Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *J. Vis.* 2, 105–120. doi: 10.1167/2.1.7
- Toga, A. W., Thompson, P. M., and Sowell, E. R. (2006). Mapping brain maturation. *Trends Neurosci.* 29, 148–159. doi: 10.1016/j.tins.2006.01.007
- Yuodelis, C., and Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Res.* 26, 847–855. doi: 10.1016/0042-6989(86)90143-4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 January 2014; accepted: 19 August 2014; published online: 08 September 2014.

Citation: Jeon ST, Maurer D and Lewis TL (2014) Developmental mechanisms underlying improved contrast thresholds for discriminations of orientation signals embedded in noise. *Front. Psychol.* 5:977. doi: 10.3389/fpsyg.2014.00977

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Jeon, Maurer and Lewis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Critical band masking reveals the effects of optical distortions on the channel mediating letter identification

Laura K. Young<sup>1,2\*</sup> and Hannah E. Smithson<sup>1</sup>

<sup>1</sup> Department of Experimental Psychology, University of Oxford, Oxford, UK

<sup>2</sup> Centre for Advanced Instrumentation, Department of Physics, Durham University, Durham, UK

## Edited by:

Rémy Allard, Université Pierre et Marie Curie, France

## Reviewed by:

Thomas S. A. Wallis, The University of Tübingen, Germany

J. Jason McAnany, University of Illinois at Chicago, USA

## \*Correspondence:

Laura K. Young, Centre for Advanced Instrumentation, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK  
e-mail: laura.young@durham.ac.uk

There is evidence that letter identification is mediated by only a narrow band of spatial frequencies and that the center frequency of the neural channel thought to underlie this selectivity is related to the size of the letters. When letters are spatially filtered (at a fixed size) the channel tuning characteristics change according to the properties of the spatial filter (Majaj et al., 2002). Optical aberrations in the eye act to spatially filter the image formed on the retina—their effect is generally to attenuate high frequencies more than low frequencies but often in a non-monotonic way. We might expect the change in the spatial frequency spectrum caused by the aberration to predict the shift in channel tuning observed for aberrated letters. We show that this is not the case. We used critical-band masking to estimate channel-tuning in the presence of three types of aberration—defocus, coma and secondary astigmatism. We found that the maximum masking was shifted to lower frequencies in the presence of an aberration and that this result was not simply predicted by the spatial-frequency-dependent degradation in image quality, assessed via metrics that have previously been shown to correlate well with performance loss in the presence of an aberration. We show that if image quality effects are taken into account (using visual Strehl metrics), the neural channel required to model the data is shifted to lower frequencies compared to the control (no-aberration) condition. Additionally, we show that when spurious resolution (caused by  $\pi$  phase shifts in the optical transfer function) in the image is masked, the channel tuning properties for aberrated letters are affected, suggesting that there may be interference between visual channels. Even in the presence of simulated aberrations, whose properties change from trial-to-trial, observers exhibit flexibility in selecting the spatial frequencies that support letter identification.

**Keywords: optical distortions, critical band masking, ocular aberrations, letter identification, spatial frequency channels, visual Strehl ratio**

## 1. INTRODUCTION

In 1994 Solomon and Pelli used critical band masking to show that, despite letters being broadband stimuli, their identification is mediated by a single narrow band of spatial frequencies. Since their ideal observer model (based on the requirement to discriminate differences between letters) exhibited low-pass filtering characteristics, rather than bandpass characteristics as derived from the performance of human observers, it was suggested that the low-frequency fall-off of the human-derived filter represents a visual constraint upon letter identification. This account is in accordance with human observers' inability to identify severely low-pass filtered letters, such as those with optical blur. Similar results have also been found by other authors (Ginsburg, 1980; Parish and Sperling, 1991; Alexander et al., 1994; Chung et al., 2002a,b; Majaj et al., 2002; Oruç and Landy, 2009).

Majaj et al. (2002) further suggested that the center frequency of the band mediating letter identification was driven by the spatial frequencies available in the signal. In the presence of added, filtered visual noise observers persisted in using the same spatial frequency channel to identify letters rather than shifting channels

to avoid the masking effects of the noise. Furthermore, when letters were filtered with a Gaussian bandpass (on a log-frequency scale) filter, the center frequency of the band mediating their identification scaled, although less than proportionally, with the center frequency of the filter. In addition to shifting the visual channel in response to filtering of the stimulus, Oruç and Landy (2009) also suggested that, when the masking noise that is added to the stimulus dominates the equivalent noise that is associated with the contrast sensitivity function of human observers (which describes how spatial frequencies are transmitted by the visual system), it is possible for an observer to switch visual channels, although not necessarily optimally.

Under natural viewing, images formed on the retina are affected by the optical quality of the eye and the distortion introduced is equivalent to filtering that image. In this paper we aim to quantify the effect that reduced image quality has on the mechanism of letter identification by considering the interaction between the spatial frequency filtering effects of optical aberrations and the spatial frequency demands of a letter identification task.

The optical quality of the human eye can be characterized by its optical transfer function (OTF), which quantifies the phase and contrast with which different spatial frequency components are transmitted. For an optically perfect eye, image quality is limited only by diffraction and the OTF is a linearly decreasing real-valued function with a cut-off frequency determined by the pupil diameter and the wavelength of light. However, real eyes are composed of imperfect optical components that introduce aberrations and these distort the wavefront of the incident light and blur the image formed on the retina. The wavefront error can now be routinely measured *in vivo* using a Shack-Hartmann aberrometer to quantify both the low order aberrations, such as defocus and astigmatism, as well as the higher-order aberrations that cannot easily be compensated with current vision correction aids.

The filtering properties of an aberration (quantified by its OTF) are different from the bandpass filters that Majaj et al. (2002) applied to letters in two respects. Firstly, the OTFs of real optical aberrations are not typically bandpass in nature but are non-monotonic and tend to attenuate high frequencies more than low frequencies. The second is that the OTFs of real optical aberrations (as opposed to Gaussian blur, for example) can be complex-valued functions, indicating spatial-frequency-dependent phase changes (quantified by the phase transfer function, PTF) in addition to spatial-frequency-dependent contrast changes (quantified by the modulation transfer function, MTF). These phase changes are an important consideration as they can have a significant impact on the spatial forms in an image. A  $\pi$  phase change, one that changes the polarity of the contrast at a particular spatial frequency, can be particularly disruptive to object recognition (Ravikumar et al., 2010) as it creates spurious resolution, which introduces additional contours. If, as suggested by Majaj et al. (2002), the visual filter mediating an identification task is selected from the signal, introduction of additional contours could bias channel selection to a sub-optimal band of frequencies.

Considering these points the following question arises—is the spatial frequency band mediating the task altered in the presence of an aberration? In this paper we use critical band masking to estimate the center frequency of the channel mediating letter identification and we measure how this is altered by three different types of aberration—defocus, coma and secondary astigmatism. We additionally mask some of the spurious resolution present in the images to look for changes in channel selection. Masking is achieved by adding (ideal) bandpass filtered noise to the stimuli and the contrast threshold elevation (from the no noise condition) for letter identification is measured for different noise center frequencies, giving a response profile. For each condition, the center frequency of the visual filter mediating the task is derived from this response profile.

There has been increasing interest in the relationship between the higher-order aberrations in the eye and visual performance. It has been shown that higher-order aberrations are detrimental to visual performance and that the reduction in performance varies between the types of aberration and their amplitudes (Applegate et al., 2002, 2003; Chen et al., 2005; Rocha et al., 2007; Zhao et al., 2009; Cheng et al., 2010; Rouger et al., 2010; Young et al.,

2013b). Furthermore, higher-order aberrations can additionally affect higher-level visual tasks such as reading (Young et al., 2011), facial recognition (Ravikumar et al., 2010; Sawides et al., 2010) and viewing natural images (Sawides et al., 2010). It is clear that the effects of these aberrations vary between visual tasks (Pepose and Applegate, 2005) and therefore it is likely that the effects of spatial-frequency-dependent changes in the stimulus depend on the spatial frequency requirements of the task. Indeed we have shown that, at least in the case of letter-based tasks, visual performance is better predicted by a model that incorporates the spatial frequencies used by the visual system for a particular task, in addition to considering image-based changes (Young et al., 2013a). Aberrations that have a strong effect on the spatial frequencies that mediate letter identification are likely to degrade performance more than aberrations that have a strong effect on task-irrelevant frequencies. Our approach to date has been to assume that the visual channel mediating letter identification is invariant under changes in aberration type and magnitude. One of the motivations of the current study is to consider the interaction between image properties and the neural channel mediating letter identification. Specification of the eye's optics permits calculation of the retinal images formed from letter-stimuli and previous work provides estimates of the neural channels mediating letter identification but, here, we consider the interaction between these two determinants of letter-identification performance.

In this paper we report the results of two analyses that aim to quantify the optical effects and predict the channel-based effects. The first analysis employs a template-matching model, based on maximum values of cross-correlations between letters, which quantifies the similarity (“confusability”) of letters for a particular aberration. We use this analysis to show the spatial frequency demands of the task based purely on the stimulus and the result is consistent with the findings of Solomon and Pelli (1994), even in the presence of an aberration. The second analysis is based on a visual Strehl metric for predicting visual performance from a measure of the optical quality of the eye. Visual Strehl metrics usually calculate the ratio of the sum under the OTF weighted by the human neural contrast sensitivity function (NCSF) to that same weighted sum for a diffraction limited system (Thibos et al., 2004). In our modified version of the visual Strehl ratio (Young et al., 2013a) we weight the OTF by the spatial frequency band mediating the task, which for sharp letters we assumed to be a Gaussian profile (in log-frequency space) with a center frequency of 3 cycles per letter (which subtended  $1^\circ$  in our experiment) and a bandwidth of 1 octave (consistent with the findings of Solomon and Pelli, 1994). Here we report visual Strehl ratios calculated using the magnitude of the OTF, termed the VSMTF, and visual Strehl ratios explicitly incorporating phase via a multiplicative combination of the MTF and PTF, which we have termed the  $VS_{combined}$  (Young et al., 2013b). We additionally calculated the VSOTF, which uses the real part of the OTF, but this gave very similar results to the VSMTF.

We have previously used both template matching and visual Strehl analyses to successfully predict the increase in contrast threshold (from the no aberration condition) for letter recognition in the presence of the three types of aberration under

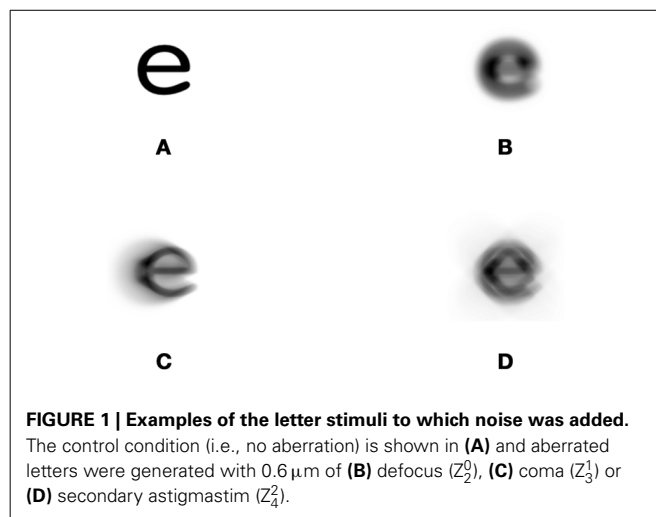
investigation. For letters presented with an aberration, and in the absence of a noise mask, the increases in contrast threshold (from the no aberration condition) quantify the performance loss over the entire frequency spectrum of those letters. Since aberrations cause spatial-frequency-dependent modifications to the stimulus it is not unreasonable to expect that the increases in contrast threshold may be spatial-frequency dependent. In this study we use critical-band masking to limit the spatial frequencies available to the observer by masking a band of frequencies with noise. For letters presented with an aberration, and in the presence of noise, the increase in contrast threshold (from the no aberration condition) should be related to the contrast loss induced by the aberration only within the spatial frequency range that remains available to the observer. For frequencies at which the aberration has little effect we would expect the increase in contrast threshold to be small. It is therefore entirely possible that the shape of the response profiles that we measure, and consequently the center frequencies derived from them, can be accounted for by considering the spatial-frequency dependent filtering of the aberrations. In this paper, by introducing additional filtering steps to our model to represent the masking effects of the visual noise, we make a direct comparison between the predicted performance and our observers' performance. We suggest that comparisons between the response profiles derived from our model and those derived from human observers should separate signal-based effects (due to image quality degradation) from observer-dependent effects (due to changes in the visual channel). Effects that are not captured by the model imply additional adaptive visual behaviors on behalf of the observer, which themselves have implications for the development of more effective models.

Finally, we recalculate the visual Strehl ratio with an additional step to optimize the center frequency of the standard Gaussian weighting that we use to represent the visual filter mediating letter identification, in order to find the best fit to the observer-derived data. Assuming that the visual Strehl metric effectively predicts performance loss due to image quality degradation, optimizing the Gaussian weighting (representing the visual channel) should capture observer-dependent effects and indicate the channel center frequency that gives rise to the contrast threshold elevations that we measure.

## 2. MATERIALS AND METHODS

### 2.1. LETTERS

In a previous experiment we tested the effects of different amplitudes (0.5, 0.6, 0.7, 0.8, and 0.9  $\mu\text{m}$  rms) of three types aberration (defocus, coma and secondary astigmatism) on observers' contrast thresholds for the identification of  $1^\circ$  letters (Young et al., 2013b). In the current experiment we have chosen to study the same three types of aberration but at a single amplitude—0.6  $\mu\text{m}$ —at which observers showed a difference in performance between these three types of aberration. This amplitude corresponds to 2.7 D of equivalent defocus over a 2.5 mm diameter pupil. As in our previous experiment, single lower-case letter images were produced as black text on a light background (in this case a gray value of 0.5 was used whereas previously it had been 1.0) using Courier font. Images of aberrated letters, such as those



shown in **Figure 1**, were generated using custom-written Python code that performed a convolution with the appropriate point spread function (PSF). We have made this code available to the community (see the Supplementary Materials).

As in our previous experiment, an unaberrated letter size of  $1^\circ$  of visual angle, corresponding to a Snellen acuity of 20/240, which is equivalent to 14 mm or 40 pt font at a typical reading distance of 40 cm, was chosen so that letter identification would be limited by the aberration and not by our observers' acuity limits. It is useful to consider how data collected at only one letter-size and one aberration amplitude might in principle generalize to other conditions. In the case of a bandpass filter applied to the letters, the center frequency of the filter is defined in cycles per letter so that when the size of a letter is changed the filtering effects remain consistent. However, whilst the OTF of an aberration can be specified as a function of frequency expressed in cycles per letter to give a consistent effect on a smaller letter, the result is not necessarily meaningful. The frequency scaling of real optical aberrations is determined by the wavelength of light and the diameter of the pupil, not by any simple combination of the size of the letters and the amplitude of the aberration. Changing the amplitude of the aberration does not simply re-scale the OTF along the frequency axis, it also changes its shape. In a previous experiment we compared the effects of an aberration at two different letter sizes (Young et al., 2013b). We used a normalized cross-correlation to find the amplitude of aberration that, when applied to a small letter, gave a similar stimulus appearance to a higher amplitude of aberration applied to a larger letter. Although the letters used in our experiments were large, and the amplitude of aberration was also correspondingly large, we know from our previous analysis that these stimuli are similar (correlation  $> 0.98$ ) to  $0.25^\circ$  letters (corresponding to a Snellen acuity of 20/60, which is equivalent to 3.5 mm or 10 pt font at a typical reading distance of 40 cm) with an aberration amplitude of 0.25  $\mu\text{m}$  rms. The aberration amplitudes found in the normal population are typically around 0.1  $\mu\text{m}$  for horizontal coma and 0.05  $\mu\text{m}$  for secondary astigmatism (Porter et al., 2001) though higher amplitudes are found in

damaged or diseased eyes, such as those with keratoconus. Thus, the aberration amplitude we have used for large letters is high compared to the amplitudes found in the normal population, but the equivalent amplitude for small letters (at sizes typically encountered when reading text) is much closer to those normally found.

Majaj et al. (2002) showed that for filtered letters, changing the size of the letter alone gave a proportional relationship between the center frequency of the channel mediating letter identification and the size of the letter. We might expect that using a smaller letter size in our experiment would result in a proportional shift in the channel frequency that we measure for our observers. For unfiltered letters, as in the no aberration condition, the center frequency could be determined based on the stroke frequency of the letters, which they defined as the number of lines crossed by a horizontal slice through a letter, divided by the letter width, and averaged over all letters.

$$\frac{f_{\text{channel}}}{10 \text{ cycles/degree}} = \left( \frac{f_{\text{stroke}}}{10 \text{ cycles/degree}} \right)^{2/3}. \quad (1)$$

The stroke frequency of the letters we used in the current experiment (1° Courier font letters) is 1.57 strokes per degree and we therefore expect the center frequency in the control condition to be 2.91 cycles per letter.

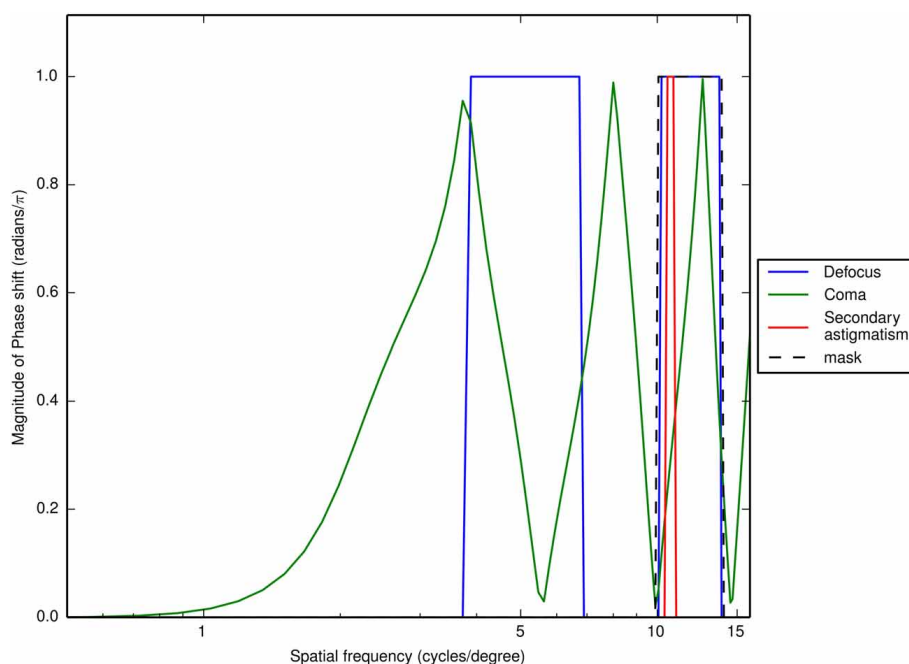
## 2.2. NOISE

White noise samples were generated by using an array of pixel values (each 0.75 arcmin pixel contributed an individual noise

check) that were sampled from a zero-mean Gaussian distribution with a standard deviation (rms contrast in this case) of 0.15 and truncated at two standard deviations. These white noise samples were bandpass filtered according to two classes of noise. The first part of the experiment was aimed at finding the center frequency of the channel mediating the identification of the aberrated letters. For this, noise samples were filtered with a one-octave-wide ideal bandpass filter centered at 1, 2, 3, 4, 6, 9, 11, or 14 cycles per degree. This was repeated with an additional bandpass filter to mask some of the spurious resolution. This additional filter was centered at 11 cycles per degree with a bandwidth of 0.5 octaves, determined by examining the OTFs of the three aberrations (see **Figure 2**). The final three conditions (9, 11 and 14 cycles per degree) were omitted when the additional mask was used as the noise-bands overlap. The entire noise field spanned 4° of visual angle and intensity values were scaled to a constant peak-to-valley contrast of 0.5 with a mean gray value of 0.5 (the same as the background of the letters).

## 2.3. STIMULUS DISPLAY

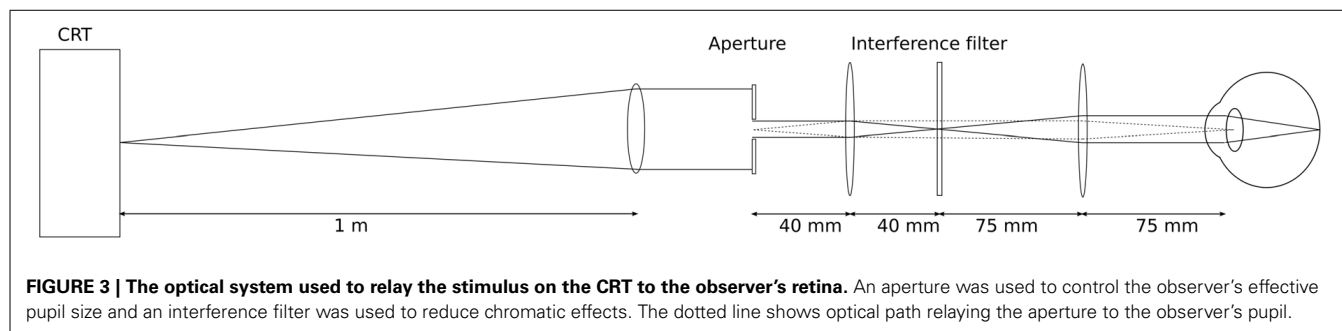
In these experiments blurred stimuli were created computationally via a convolution of a PSF with an image of a letter (see the Supplementary Materials). The resulting stimuli represent the image that would be formed on the retina by an eye with the specified PSF. To ensure that this was indeed the image formed on the retina of our observers it was necessary to consider the effects of aberrations introduced by the observer's eye and also to compensate for the effects of the display and the optics that relay the stimulus to the observer's eye (see **Figure 3**).



**FIGURE 2 | The magnitude of the phase transfer functions for 0.6 μm of defocus ( $Z_2^0$ ), coma ( $Z_3^1$ ) and secondary astigmatism ( $Z_4^2$ ). An additional band of noise was added centered at 11 cycles per degree with a bandwidth**

of 0.5 octaves (indicated by the dashed line) to coincide with spurious resolution created by defocus and secondary astigmatism. The phase shifts for coma are wrapped.





The resolution of an image formed on the retina is limited by the optical quality of the eye and the fundamental limit is that imposed by diffraction. Considering only the effects of diffraction it is best to use a large pupil to obtain the highest resolution images, however aberrations in the eye tend to increase with increasing pupil size. The optimal pupil size for lateral resolution has been shown to be 2.5 mm (Campbell and Green, 1965; Donnelly and Roorda, 2003), which is a diameter that has previously been used when simulating the effects of higher-order aberrations (see Cheng et al., 2010; Young et al., 2013b, for example). Additionally, we have measured the aberrations in our observers' eyes using a Zywave aberrometer and can confirm that over a 2.5 mm pupil they are close to diffraction-limited. Chromatic aberrations were avoided by using a narrowband interference filter centered at a wavelength of 550 nm.

The display and the optical system change the contrast of spatial frequencies in the image, quantified by their individual MTFs. The MTF of the entire optical system, including the display and the aperture, was measured via the slanted edge method (Estibean and Magnan, 2004) using a camera that had been calibrated using the same technique in conjunction with an ISO 12233 test chart. This measure of the MTF of the optical system was used to pre-compensate the images for the contrast changes caused by the optical system.

The pre-compensated images were displayed on a CRT (Sony Trinitron, 1024 × 768 resolution) display using a Cambridge Research Systems VSG stimulus generator (VSG2/5) and the CRS Matlab toolboxes. To account for the intensity non-linearity in the display a gamma correction was applied to the stimuli using a look-up table, which was specified to maintain a resolution of 8 bits per gun for all stimulus contrasts, selected from  $2^{12}$  available gray levels across the full intensity range. The letter image and the noise image were combined by temporally interleaving frames at a rate of 100 Hz. The mean luminance of the monitor measured through the optical system was  $7.75 \text{ cd m}^{-2}$ .

Due to space constraints the image on the monitor had to be demagnified to maintain an acceptable sampling rate at the retina. In this arrangement, a single pixel on the display spanned 0.75 arc min on the retina, giving a sampling frequency of 80 pixels per degree. To prevent aliasing all stimuli were digitally low-pass filtered with a cut-off frequency of 40 cycles per degree. An aperture was used to artificially stop the pupil down to 2.5 mm and this was relayed to the eye's pupil using the optical system shown in Figure 3. This system produced a magnification factor of two between the artificial pupil and the observer's pupil,

so an artificial pupil diameter of 1.25 mm was used. The cut-off frequency,  $f_{\text{cut-off}}$ , of an optical system is defined by:

$$f_{\text{cut-off}} = \frac{D}{\lambda}, \quad (2)$$

where  $D$  is the diameter of the aperture and  $\lambda$  is the wavelength of light. The cut-off frequency in the intermediate focus, resulting from the aperture, was 40 cycles per degree. After demagnification this corresponded to a frequency of 80 cycles per degree at the retina, which was well above the cut-off frequency of the digital low-pass filter.

## 2.4. PROCEDURE

The study received ethical approval from the Medical Sciences Division (MSD) Interdivisional Research Ethics Committee (IDREC) which operates under the Central University Research Ethics Committee (CUREC) at the University of Oxford. Informed consent was obtained from all observers. Three observers, two aged 28 and one aged 37, took part in the experiment. Two of the observers required refractive correction and so wore contact lenses. Observers were aligned to the instrument prior to beginning the experiment and were held in position by a chin rest. A separate rest was carefully positioned in front of the eye that observers could comfortably rest their cheek and brow bones against, allowing them to re-align themselves. Stimuli were displayed monocularly for 200 ms after which observers responded via a keyboard and audio feedback was then given.

For each trial, a letter was chosen at random and the probability of selection was weighted by the frequency counts of letters in the English language (Jones and Mewhort, 2004) in order to simulate natural reading conditions. The contrast of the noise was kept constant and Weber contrast thresholds for the letters were measured using the ML-PEST algorithm implemented with the Matlab Palamedes Toolbox (Prins and Kingdom, 2009). The algorithm converged to the threshold corresponding to 64% correct. Individual staircases for the experimental conditions were interleaved with each running for 20 trials per experimental condition and observers completed five sessions.

## 2.5. PREDICTION METRICS

To investigate the effect of an aberration on image quality we used two types of metric, one based on template matching and one based on the visual Strehl ratio. In both cases we specifically considered the frequency-dependent changes, to consider the effect of the aberration on channel selection.

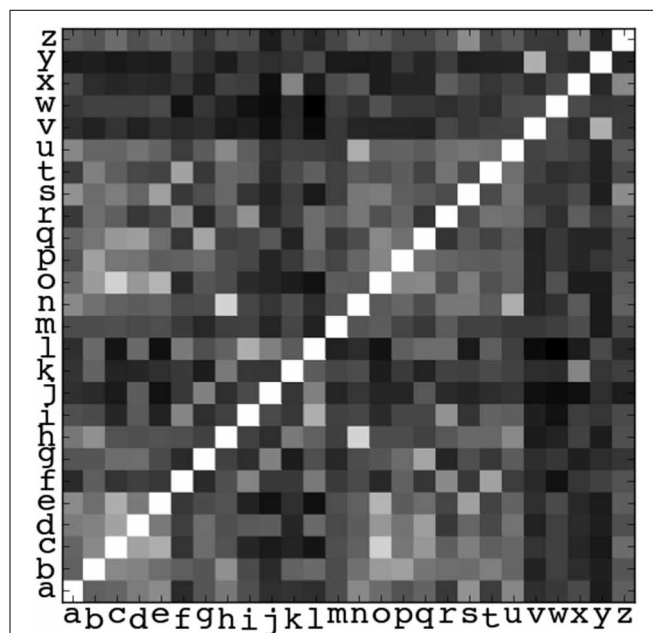


### 2.5.1. Template matching

Cross-correlation-based template matching models have been shown to have a high correlation with acuity measures (Watson and Ahumada, 2008, 2012, for example). We used a similar technique that we have previously shown to correlate well with empirical measures of performance in letter-based tasks (Young et al., 2011, 2013a,b). For the current experiment we added an additional filtering step to account for the masking effects of the noise. This technique made pair-wise comparisons between letters via a cross-correlation, as described in the following steps, which were repeated for each type of aberration and noise pass-band: (i) all of the letters of the alphabet were individually notch-filtered to remove the spatial frequencies that would be masked by noise in the experiment, (ii) the maximum of the cross-correlation between pairs of filtered letter images (one filtered, aberrated letter and one filtered, unaberrated letter) formed a 26-by-26 matrix, an example of which is given in **Figure 4**, (iii) the confusion matrix was normalized to one along the diagonal, (iv) the columns of the confusion matrix were weighted according to the frequency with which letters appeared in the experiment, (v) the average value of the matrix was used as a measure of confusability between letters, (vi) confusability values were scaled such that a value of zero means that the only overlap is between a letter and its (unaberrated) template (i.e., the raw, unweighted correlation matrix is the identity matrix), and a confusability value of one means that all aberrated letters overlap with the unaberrated template by the same amount as the aberrated version of the template letter. The confusability value for unaberrated, unfiltered letters is 0.6. The confusion analysis was performed for comparisons between the aberrated letters and the unaberrated letter templates, as well as between pairs of aberrated letters. Similar results were obtained in both cases. Here we report only the comparisons between aberrated letters and the unaberrated letter templates.

### 2.5.2. Visual Strehl metrics

Using two types of visual Strehl metric (VSM TF and  $VS_{combined}$ ) we modeled the effect of the change of image quality on observers' performance. The visual Strehl ratio is a measure of (neurally weighted) relative image quality, quantifying the ratio of the sum under the OTF of an aberrated optical system to that of a diffraction limited one. In the traditional visual Strehl ratio (Thibos et al., 2004) the OTF is weighted by the human neural contrast sensitivity function, which attenuates high and very low spatial frequencies. We recently modified this metric such that the OTF is weighted by the neural filter that mediates the task (which for letter identification we had assumed to be a Gaussian function with a mean of 3 cycles per letter and a bandwidth of 1 octave) to give improved predictions of performance (Young et al., 2013a). In this paper we perform the same calculations but with two additional modifications. Firstly, for each observer, instead of using a Gaussian weighting,  $LB(f_x, f_y)$ , with a mean of 3 cycles per letter we use a mean equal to the center frequency derived from the observer's performance in the control condition. Secondly, as for the template matching model, we introduce an additional notch filter,  $NF(f_x, f_y)$ , to account for the masking effects of the noise. The equations representing the VSMF and  $VS_{combined}$  are



**FIGURE 4 |** An example confusion matrix indicating the maximum values of cross-correlations between letter images, in this case for the no aberration condition with no notch filter. Gray values indicate values ranging from zero (black) to one (white). The matrix is normalized such that the values along the diagonal are equal to one. The confusability value associated with this matrix is 0.6.

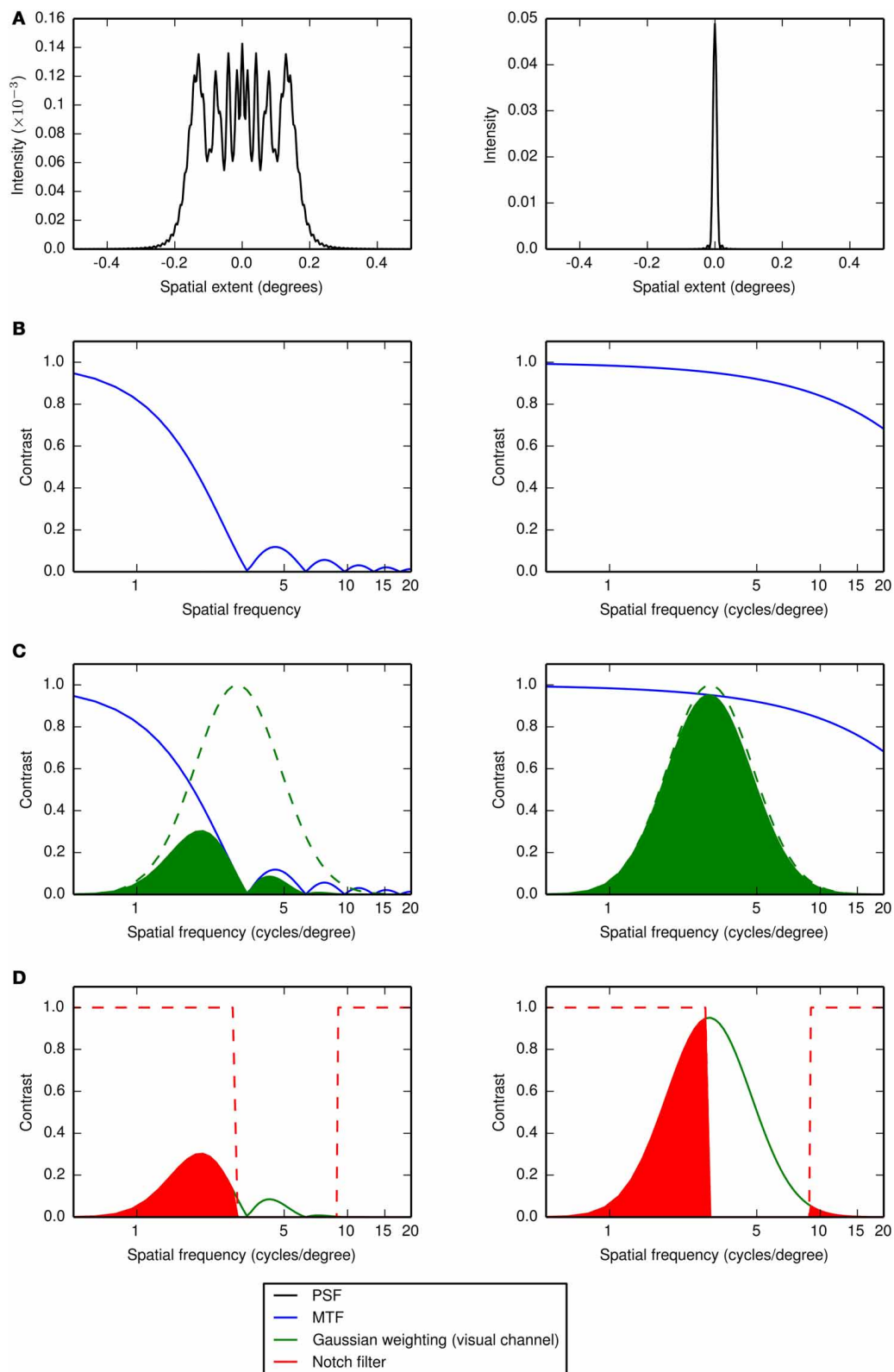
$$VSM TF = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} MTF(f_x, f_y) \cdot LB(f_x, f_y) \cdot NF(f_x, f_y) df_x df_y}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} MTF_{DL}(f_x, f_y) \cdot LB(f_x, f_y) \cdot NF(f_x, f_y) df_x df_y}, \quad (3)$$

and

$$VS_{combined} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} MTF(f_x, f_y) \cdot \left(1 - \frac{|PTF(f_x, f_y)|}{\pi}\right) \cdot LB(f_x, f_y) \cdot NF(f_x, f_y) df_x df_y}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} MTF(f_x, f_y) \cdot LB(f_x, f_y) \cdot NF(f_x, f_y) df_x df_y}, \quad (4)$$

where  $MTF(f_x, f_y)$  is the MTF of the aberrated PSF,  $MTF_{DL}(f_x, f_y)$  is the diffraction-limited MTF and  $PTF(f_x, f_y)$  is the PTF (in the range  $-\pi$  to  $\pi$ ) of the aberrated PSF. A summary of the method for calculating the visual Strehl ratio is given in **Figure 5** and further details can be found in our previous paper Young et al. (2013a).

As we have already shown that modified visual Strehl metrics are a good predictors of the increase in contrast threshold for letter identification from the no aberration condition, we would expect these results to closely match observers' change in noise-masked performance if observers continued to use the same band of spatial frequencies for each type of aberration. If this is the case we can assume that the change in the response profile (associated with an aberration) is caused by a reduction in image quality



**FIGURE 5 | The steps used to calculate the visual Strehl ratio (VSMTF in this example).** The numerator in equation 3 is calculated as follows: For a particular aberration (shown on the left half of the figure) **(A)** the PSF is computed from the wavefront error associated with the aberration, **(B)** the

MTF is computed as the Fourier transform of the PSF, **(C)** the MTF is weighted by a Gaussian function (in log frequency space and with mean equal to an observer's channel center frequency in the control condition)

(Continued)

**FIGURE 5 | Continued**

representing the visual channel mediating letter identification and **(D)** the result (green shaded area) is then additionally weighted by a notch filter that removes the spatial frequencies masked by the passband of noise. The result of this additional weighting (red shaded area) is then summed. The denominator in equation 3 is calculated by repeating steps **(A–D)** for a diffraction-limited PSF, as shown on the right half of the figure. The visual

Strehl ratio is then the sum for the aberration divided by the sum for the diffraction-limited case and the result is weighted by the total signal power in the notch filter. A similar calculation is performed for the  $VS_{combined}$  (see Equation 4) except that in the numerator the MTF is weighted by the PTF, normalized between zero (at a phase shift of  $\pi$ ) and one (at a phase shift of zero) and in the denominator the diffraction-limited MTF is replaced by the MTF for the aberration (Young et al., 2013a).

alone. If this is not the case, we wish to estimate which band of spatial frequencies our observers are using.

To estimate shifts in the putative neural channel underlying performance over and above changes imposed by the frequency-dependent changes in image quality, we additionally determine the center frequency of the Gaussian weighting (letter band, *LB*) with which to weight the OTF (as step **Figure 5C**), so as to minimize the sum of the squared differences between the visual Strehl ratio and the observer-derived increase in threshold (from the no aberration condition), within each particular noise band.

### 3. RESULTS

#### 3.1. THRESHOLD ELEVATION

For individual observers, threshold energy values,  $E$ , were calculated from the contrast threshold measure for each experimental condition and these were averaged over five sessions. These thresholds could be a result of intrinsic visual noise as well as the noise we have added. Therefore, we quote threshold elevations,  $E^+ = E - E_0$ , where  $E_0$  is the average threshold energy with no external noise. We also quote threshold signal-to-noise ratios, *SNR*, which are calculated as:

$$SNR = \frac{E^+}{N}, \quad (5)$$

where  $N$  is the noise power spectral density (Pelli and Farrel, 1999).

Thresholds were obtained as a function of the center frequency of the one-octave ideal bandpass noise. The center frequency of the channel was estimated from the location of the maximum threshold elevation determined by fitting a Gaussian function (on a log-frequency scale) to the threshold signal-to-noise ratio and extracting the mean value. Thresholds were additionally obtained in the presence of an additional passband of noise designed to mask spurious resolution. The threshold elevations and signal-to-noise ratios are given in **Figure 6** (simple bandpass mask) and **Figure 7** (bandpass mask plus additional band to mask spurious resolution) and the center frequencies are summarized in **Figures 10A,B**.

Fitting Gaussian profiles to a measure of contrast threshold elevation gave an average center frequency (across observers) of 2.3 cycles per letter (standard error,  $SE = 0.1$  cycles per letter) for defocus, 2.7 cycles per letter ( $SE = 0.1$  cycles per letter) for coma, 2.3 cycles per letter ( $SE = 0.1$  cycles per letter) for secondary astigmatism and 3.6 cycles per letter ( $SE = 0.1$  cycles per letter) in the control condition. With the additional spurious resolution mask the average center frequencies were 1.7 cycles per letter (standard error,  $SE = 0.1$  cycles per letter) for defocus, 2.6 cycles per letter ( $SE = 0.2$  cycles per letter) for coma, 2.2 cycles

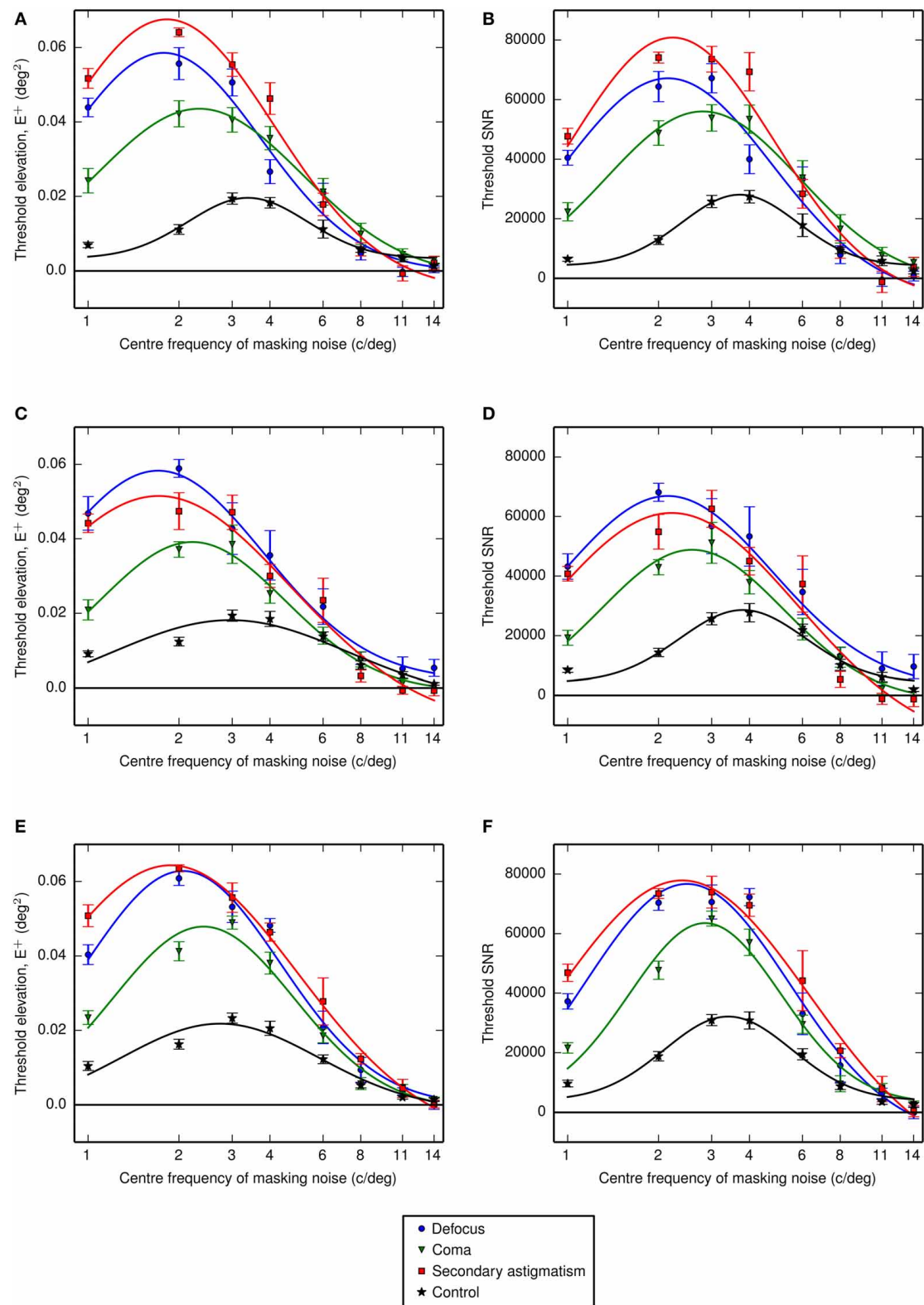
per letter ( $SE = 0.2$  cycles per letter) for secondary astigmatism and 3.2 cycles per letter ( $SE = 0.2$  cycles per letter) in the control condition.

These fitted center frequencies revealed a shift from the no aberration condition of between  $-0.45$  and  $-0.82$  octaves for defocus, between  $-0.25$  and  $-0.56$  octaves for coma and between  $-0.51$  and  $-0.78$  octaves for secondary astigmatism. The threshold signal-to-noise ratios show a consistent increase in threshold with the additional spurious resolution mask in the presence of an aberration (**Figure 7**). However, the spurious resolution mask only produced a significant shift (as determined by the 95% confidence limits) in center frequency for defocus for observers RJL and HES, and this was to lower frequencies. For observers LKY and HES, there is little difference between the control condition with the spurious resolution mask and the control condition without it, suggesting that this additional mask is having a negligible effect on performance. However, observer RJL exhibited a significant shift in the channel frequency for the control condition.

#### 3.2. PREDICTING THE CHANNEL

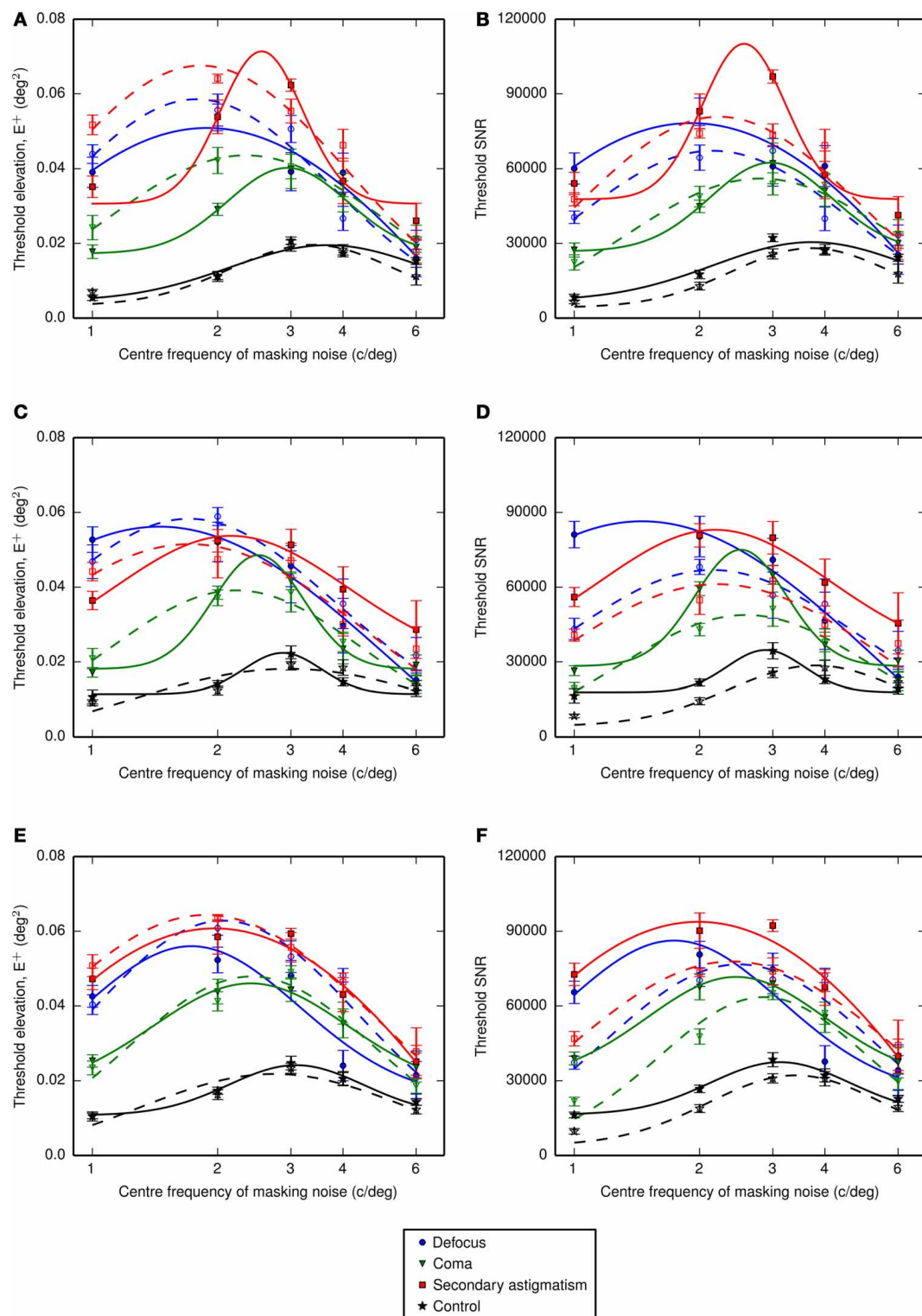
For unaberrated, unfiltered letters the confusability value is 0.60. For defocus, coma and secondary astigmatism the confusability values for unfiltered letter stimuli are 0.80, 0.78 and 0.82 respectively. Unsurprisingly, the aberrations increase confusability. But, of interest in our present experiment is how the distinguishing features that remain are distributed across the spatial frequency range. **Figure 8** shows the relationship between confusability and the center frequency of the notch filter that had been applied to the letters to represent the masking effects of the bandpass noise. A high confusability value suggests that removing the band of frequencies (via filtering in our model or via masking in the experiment) makes letters more difficult to distinguish and therefore the identification task should be more difficult and the associated contrast thresholds correspondingly larger. The results show that, based purely on the demands of a template matching task, the response should be low pass, which is in agreement with the findings of Solomon and Pelli (1994). The low-pass characteristic found with unaberrated letters is retained for aberrated letters.

To investigate the effects of a reduction in image quality caused by each aberration, we used the visual Strehl metrics described in Section 2.5.2. The visual Strehl ratios, shown in **Figure 9**, indicate the change in visual image quality associated with the addition of an aberration, within the band of frequencies we assume observers to be using for the task (based on their threshold elevations in the control condition). If a change in image quality were the only factor driving the difference in performance between aberrated and unaberrated letters, we would expect these values to correlate with the increases (from the no aberration condition)



**FIGURE 6 | Threshold elevations (above the no noise condition) for letter identification measured with noise that had been filtered using one-octave-wide (ideal) passbands at different center frequencies. (A,B)** Show the threshold energy elevations and threshold signal to noise

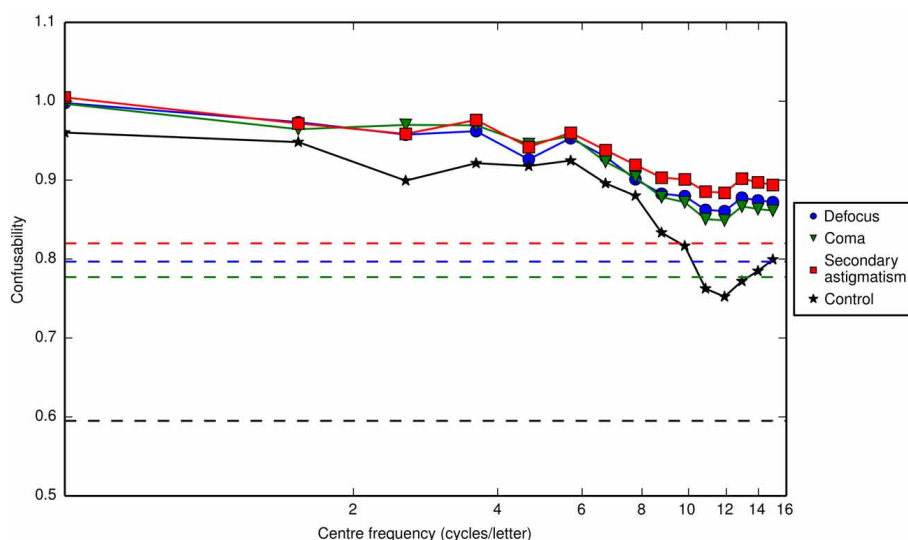
ratios for observer LKY, **(C,D)** show the equivalent results for observer RJL and **(E,F)** for observer HES. The frequency axis is shown on a  $\log_{10}$  scale and Gaussian functions were fitted to the data in log-frequency space. Error bars represent the standard error on the mean.



**FIGURE 7 |** Threshold elevations (above the no noise condition) for letter identification measured with noise that had been filtered using one-octave-wide (ideal) passbands at different center frequencies plus an additional band to mask spurious resolution in the letter images. (A,B) Show the threshold energy elevations and threshold signal to noise ratios for observer LKY, (C,D) show the

equivalent results for observer RJL and (E,F) for observer HES. The dashed lines with open symbols show the data in the absence of the spurious resolution mask, replotted from **Figure 6**. The frequency axis is shown on a  $\log_{10}$  scale and Gaussian functions were fitted to the data in log-frequency space. Error bars represent the standard error on the mean.





**FIGURE 8 | The confusability of letters as a function of the center frequency of a one octave-wide notch filter applied to the letters.**

A confusability value of zero means that the only overlap is between a letter and its (unaberrated) template (i.e., the raw, unweighted correlation matrix is the identity matrix), and a confusability value of one means that all the aberrated letters overlap to the same degree as they do with the (unaberrated) template (i.e., the raw, unweighted correlation matrix is a matrix of ones). A confusability value greater than one can be obtained, if on average, there is more overlap between an “incorrect” aberrated letter and the (unaberrated) template than there is between the “correct” aberrated

letter and its (unaberrated) template. The confusability values for unfiltered letters are 0.60 in the control condition and 0.80, 0.78 and 0.82 for defocus, coma and secondary astigmatism respectively (indicated by the dashed lines). For all three types of aberration and the control, the confusability of letters is highest for low frequency notch filters. The notch filter removes a band of frequencies from the image, which we use to simulate bandpass filtered noise masking those same frequencies. The results suggest that the image characteristics that distinguish letters are concentrated at low spatial frequencies, since removing these frequencies increases confusability the most.

in threshold SNR measured empirically. **Figure 9** shows that this is not the case.

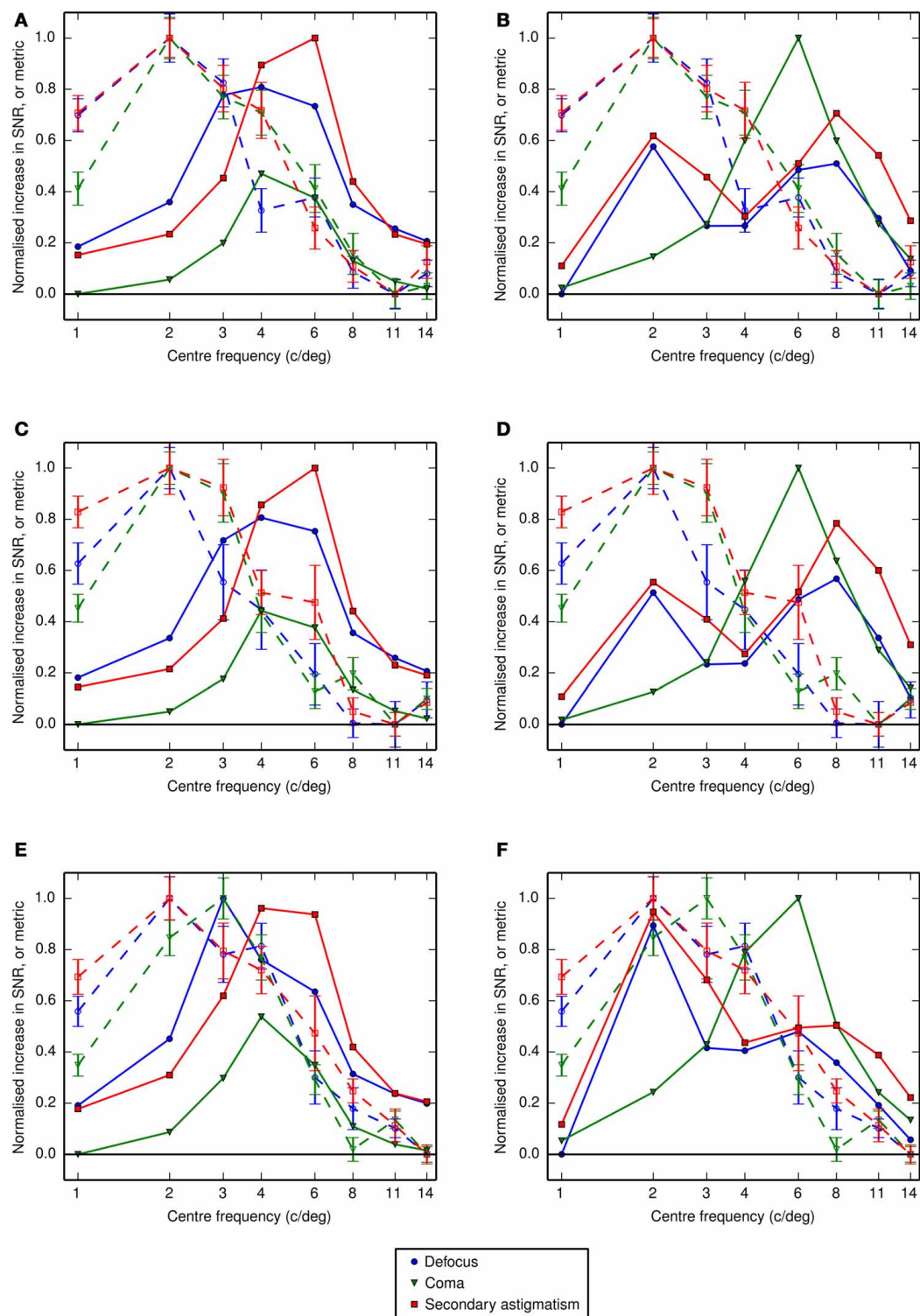
To determine the weighting that should be applied to the OTF to represent the visual filter our observers are actually using in the aberration conditions, a sliding filter (as opposed to one generated from observers’ measured center frequencies in the control condition) was used to reproduce the analysis summarized in **Figure 5**. Using this method we determined the center frequency of the sliding filter that produced visual Strehl ratios that most closely matched the observer-derived increase (from the no aberration condition) in threshold SNR. The results, given in **Figures 10C,D**, suggest that in the presence of an aberration the center frequency of the visual filter mediating letter identification is most likely lower than that for the control condition.

#### 4. DISCUSSION

When studying the effects of an optical aberration on visual performance it is important to consider not only the degradation of the image per se but specifically the loss of information in the image that is required to succeed at a visual task. Previous work has shown that visual performance is impaired in the presence of optical aberrations (Applegate et al., 2002, 2003; Chen et al., 2005; Rocha et al., 2007; Zhao et al., 2009; Cheng et al., 2010; Rouger et al., 2010; Young et al., 2013b). The reduction in visual performance is related to both the type of aberration and the amplitude of that aberration (Applegate et al., 2003, for example) and the effect is task-specific (Pepose and Applegate, 2005). Most work in this respect has focussed on letter-based tasks, most

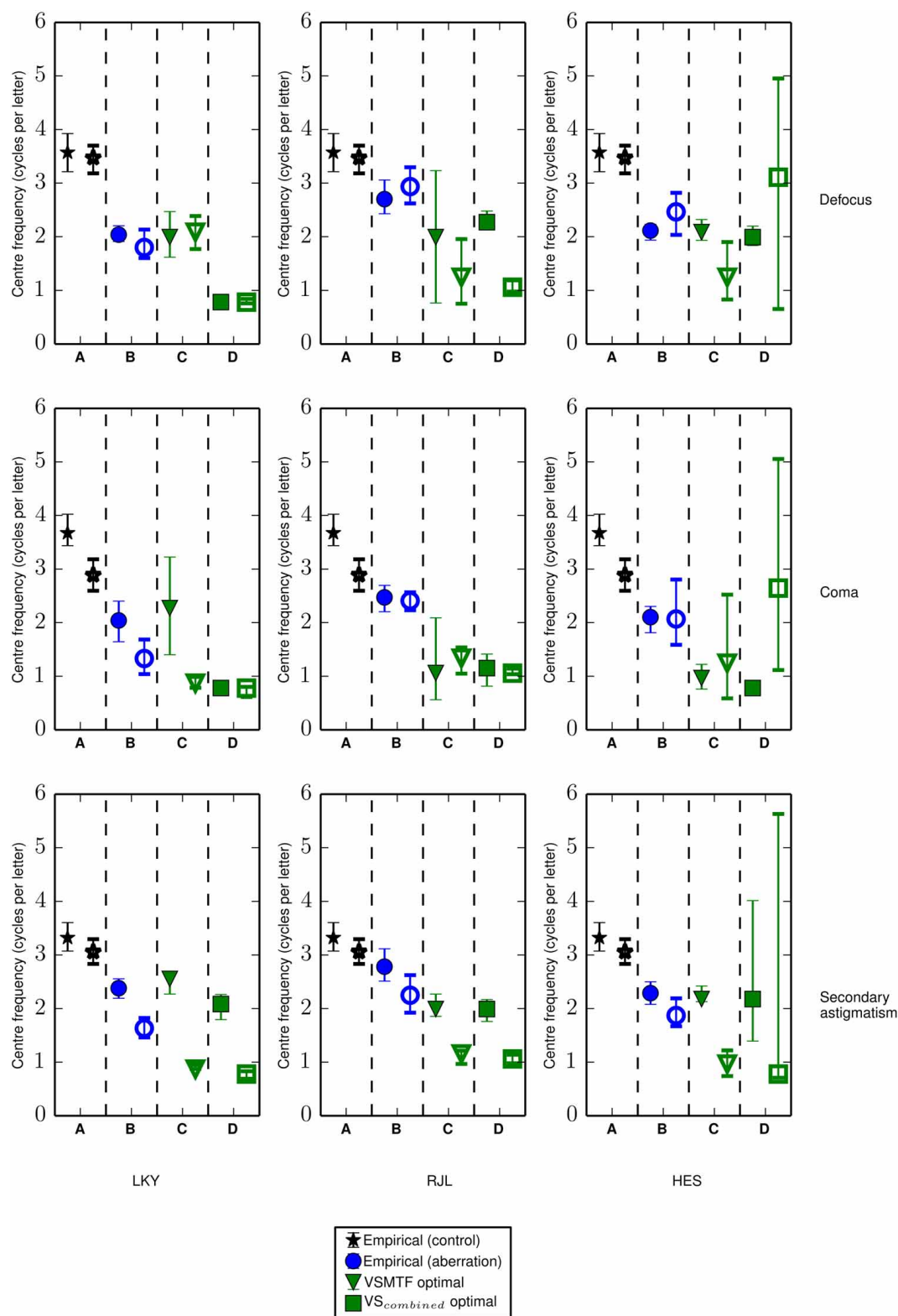
likely because they are a standard clinical optotype for assessing visual impairment and, being over-learned, they are ideal for testing object recognition. Aberrations cause spatial-frequency-dependent changes in an image and with letters being broadband stimuli we might expect that visual information is disrupted across the entire spectrum. However, any method for predicting performance that makes this assumption has the potential to incorrectly estimate the performance loss, since it will include degradation at spatial frequencies that do not ultimately mediate the task. Restricting the band of spatial frequencies over which a prediction metric is calculated to those mediating the task can mitigate this problem.

Solomon and Pelli (1994) showed that an ideal observer model predicts that the response profile, based purely on the demands of the task, should be low-pass. We additionally show that, even in the presence of aberrations, the response of an ideal observer (which we model via template matching) should also be low pass (**Figure 8**). Contrary to the predictions of their ideal observer model, Solomon and Pelli (1994) found, using critical band masking, that human observers use a single band of spatial frequencies centered at three cycles per letter with a bandwidth of one to two octaves. Further to this, Majaj et al. (2002) demonstrated that filtering a letter (at a fixed size) causes the center frequency to shift, scaling less than proportionally with the center frequency of the filter (defined in cycles per letter). Oruç and Landy (2009) also proposed that observers could switch spatial-frequency channels, but not necessarily to the optimal one. We hypothesized that real ocular aberrations, which act to spatially filter an image, could



**FIGURE 9 | (A,B)** Comparison between observer LKY's measured performance (dashed lines, open symbols) and a prediction of performance based on the visual Strehl ratio (solid lines, closed symbols) computed using (A) the VSMTF and (B) the  $VS_{combined}$ . Panels (C,D) are the equivalent data for observer RJL and (E,F) are those for observer HES. As the visual Strehl ratio is high for good image quality the data are presented as the reciprocal of the visual Strehl ratio for comparison with threshold values. The

observer-derived values reported here are the increase in threshold signal-to-noise ratio from the no aberration condition (i.e., the data presented as colored lines in the right half of **Figure 6** minus the corresponding control data, presented as black lines), and in the absence of the additional spurious resolution mask. If our observers' performance were affected only by image quality degradation, and not a shift in the putative neural channel, the dashed lines should overlap with the solid lines.



**FIGURE 10 | The channel center frequencies derived from Gaussian fits (in log-frequency space) to the response profiles or by running the visual Strehl model.** Closed symbols show center frequencies in the absence of the spurious resolution mask and open symbols show the center frequencies with the additional spurious resolution mask. Data from different

aberration types are shown in rows and data from different observers are shown in columns. Each sub-panel shows (A) the center frequency in the control condition (replotted in each row) derived from the empirical measure of threshold signal-to-noise ratio (as shown in right half of **Figures 6, 7**),

(Continued)

**FIGURE 10 | Continued**

(B) the similarly derived center frequency in the aberration condition, where the aberration type is indicated by the labels on the right of the figure, (C) the optimal center frequency for the Gaussian weighting (i.e., the putative visual channel) when calculating the VSMTF to most closely match the increase in the observer-derived threshold signal-to-noise ratio from the no aberration condition and (D) the optimal center frequency for

the Gaussian weighting (i.e., the putative visual channel) when calculating the  $VS_{combined}$  to most closely match the increase in the observer-derived threshold signal-to-noise ratio from the no aberration condition. Error bars represent the 95% confidence limits, which were calculated by bootstrapping the data for each observer and each condition, drawing randomly with replacement from the data for individual trials, repeated for 1000 simulations.

affect the spatial frequency channel mediating letter identification. Additionally, some aberrations create spurious resolution that introduces extra contours, which may drive the channel to a sub-optimal center frequency.

By using a critical-band masking technique for aberrated letter stimuli we have demonstrated measurable shifts in the channel frequency mediating their identification to lower frequencies. A shift from higher to lower spatial frequencies may be expected since aberrations generally attenuate higher frequencies but maintain contrast better at lower frequencies, yet confusability is less low pass in aberrated conditions than in the control (Figure 8). Across all observers there was a consistent trend showing that defocus and secondary astigmatism caused the largest shift in frequency band producing the peak masking effect and coma showed a smaller change.

Another possible effect could be that the bandwidth of the channel changes in the presence of an aberration if the information in the original channel is insufficient to identify the letter. We also tested the effects of changing the bandwidth of the noise by repeating the experiment with filtered noise having a varying bandwidth centered on the center frequencies determined for individual observers. These results are not presented here because they were too noisy to draw any firm conclusions from, most likely because the spatial frequency dependent contrast changes caused by these aberrations are non-monotonic. As the bandwidth increases it could potentially mask spurious resolution, or any other phase changes that may disrupt letter identification, and performance may partially improve. The results demonstrated a sigmoid shape but with additional dips that would be consistent with this hypothesis. However, as an approximation to the expected sigmoid shape it appeared that the bandwidth of the filter was not changing dramatically.

Ideally we would have compared thresholds measured in band-pass noise with those measured in notch filtered noise. This would have indicated whether observers were channel switching. While we did attempt this, any effects from the interaction between the notch filter and the aberration were lost in the noise due to having insufficient dynamic range for the noise contrast. We chose instead to look specifically at masking a secondary band of frequencies, coinciding with spurious resolution to investigate interactions between frequency channels.

If threshold performance depended only on a narrow band of frequencies, centered on 3 to 4 cycles per letter, adding the additional high-frequency mask should have no effect on performance. Our results on the other hand show a consistent increase in threshold with the additional mask (Figure 7) and in aberrated conditions there is a suggestion that the mask shifts the putative neural channel to lower frequencies (Figure 10B). Clearly there is some interaction between channels, although our results are

insufficient to describe the nature of this interaction. The effects of spurious resolution are likely to be complex, perhaps producing false positives based on contours or features that might wrongly identify a letter, or driving looking to a suboptimal channel that has relatively high contrast. However, the way in which this is affected by the mask and by the observer's familiarity with the stimuli is beyond what we have tested here.

What is not clear from the data in Figures 6, 7 is whether the threshold elevations simply represent the residual information available in the noise-masked stimulus, or whether there is an interaction between the spatial frequencies used by the observer and the nature of the aberration. If observers persist in using the same visual channel in the presence of an aberration, we should be able to predict the change in threshold simply from the change in image quality within that channel. We have chosen to use visual Strehl metrics for this analysis. Our conclusions are dependent on these metrics providing an adequate summary of image quality, which is supported by several studies showing that they work well for predicting performance on letter-based tasks in the presence of an aberration (e.g., Marsack et al. 2004; Thibos et al. 2004; Young et al. 2013b). Using the center frequencies measured for the control condition we calculated the visual Strehl ratio in each noise filter band. Figure 9 shows how the increase in threshold SNR (from the no aberration condition) of our observers should be skewed if their performance were affected only by image quality degradation in the absence of any change in the visual channel they were using. It is clear that the values predicted from the visual Strehl ratio do not match the observer-derived increases in threshold SNR and we therefore assume that the our observers are using a different band of spatial frequencies than those derived in the control condition.

The response profile derived from an observer's threshold SNR in the control condition can be used to infer the center frequency of the neural channel associated with letter recognition for an optically ideal visual system. The corresponding response profiles measured in the aberration conditions include the effects of changes in image quality and any changes in the putative neural channel supporting performance. We use a modified visual Strehl ratio to capture the changes in the image (from the no-aberration condition) and the changes in the neural channel by finding the neural filter that, when used as the weighting function in the visual Strehl calculation, gives the best match between the metric and the measured increases in threshold SNR. We consider aberration-induced changes in the putative channel mediating letter identification to be the shifts in center frequency from the no-aberration condition (in which there are no image quality effects) to the center frequency of the sliding filter that most closely captures the measured performance changes. Figures 10C,D summarizes these results. The VSMTF



reveals shifts in the putative visual channel to much lower spatial frequencies (as compared to the control condition, panel in **Figure 10A**), and this was also observed in the presence of the additional spurious resolution mask. The  $VS_{combined}$  metric produces inconsistent results across observers and this is most likely due to the non-monotonic nature of the  $VS_{combined}$  profiles. The non-monotonic profiles arise due to additional  $\pi$  phase changes at lower spatial frequencies (e.g., see **Figure 2**). Interestingly, these have a substantial effect on the metric but not on performance. Accurately modeling the consequences of phase changes for visual performance is difficult, and these results suggest the  $VS_{combined}$  metric should be further refined.

It is important to note that our observers were not adapted to these aberrations, as they might be if they were permanent feature of their vision. Therefore we cannot be sure that an observer with these aberrations occurring naturally would have a shifted center frequency with respect to the normal population. Additionally we have tested these three types of aberration in isolation whereas in a normal eye there would be combination of aberration types and amplitudes.

Our results suggest that the impact optical aberrations have on letter identification performance is not only based upon a loss of contrast or changes in phase but that there is also the potential for them to alter the neural channel selected to support letter identification. We already know that optical aberrations can have far reaching effects on visual performance as we have shown that certain types of aberration specifically affect the process of word recognition (Young et al., 2011) with uncommon words taking disproportionately longer to identify than common words in the presence of defocus or secondary astigmatism than with no aberration applied. Majaj et al. (2002) suggested that the channel mediating letter identification is selected bottom-up by the signal and our results broadly agree with this hypothesis in that observers' response profiles changed when an aberration was present. The measured change in performance was not simply predicted by the spatial-frequency-dependent change in image quality within a fixed channel, at least based on the image quality metrics we have used, suggesting that observers exhibit flexibility in the channel they select for letter identification in the presence of an aberration.

## FUNDING

We would like to acknowledge the John Fell Fund for supporting this work.

## ACKNOWLEDGMENT

We wish to thank Bruce Henning for an interesting discussion on noise masking techniques, Robert Lee for his help running the experiment and Rowan Bolton for his work with the slanted edge algorithm.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01060/abstract>

## REFERENCES

- Alexander, K. R., Xie, W., and Derlacki, D. J. (1994). Spatial-frequency characteristics of letter identification. *J. Opt. Soc. Am. A* 11, 2375–2382. doi: 10.1364/JOSAA.11.002375
- Applegate, R. A., Ballentine, C., Gross, H., Sarver, E. J., and Sarver, C. A. (2003). Visual acuity as a function of Zernike mode and level of root mean square error. *Optom. Vis. Sci.* 80, 97–105. doi: 10.1097/00006324-200302000-00005
- Applegate, R. A., Sarver, E. J., and Khemsara, V. (2002). Are all aberrations equal? *J. Refract. Surg.* 18, S556–S562.
- Campbell, F. W., and Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *J. Physiol.* 181, 576–593.
- Chen, L., Singer, B., Guirao, A., Porter, J., and Williams, D. R. (2005). Image metrics for predicting subjective image quality. *Optom. Vis. Sci.* 82, 358–369. doi: 10.1097/01.OPX.0000162647.80768.7F
- Cheng, X., Bradley, A., Ravikumar, S., and Thibos, L. N. (2010). Visual impact of Zernike and Seidel forms of monochromatic aberrations. *Optom. Vis. Sci.* 87, 300–312. doi: 10.1097/OPX.0b013e3181d95217
- Chung, S. T. L., Legge, G. E., and Tjan, B. S. (2002a). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vis. Res.* 42, 2137–2152. doi: 10.1016/S0042-6989(02)00092-5
- Chung, S. T. L., Levi, D. M., and Legge, G. E. (2002b). Spatial-frequency properties of letter identification in amblyopia. *Vis. Res.* 42, 1571–1581. doi: 10.1016/S0042-6989(02)00065-2
- Donnelly, W. J., and Roorda, A. (2003). Optimal pupil size in the human eye for axial resolution. *J. Opt. Soc. Am. A* 20, 2010–2015. doi: 10.1364/JOSAA.20.002010
- Estibeau, M., and Magnan, P. (2004). Fast mtf measurement of cmos imagers using iso 12233 slanted-edge methodology. *Proc. SPIE* 5251, 243–251. doi: 10.1117/12.513320
- Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: an explanation of how we see certain objects. *Proc. SID* 21, 219–227.
- Jones, M. N., and Mewhort, D. J. K. (2004). Case-sensitive letter and bigram frequency counts from large-scale English corpora. *Behav. Res. Methods Instrum. Comput.* 36, 388–396. doi: 10.3758/BF03195586
- Majaj, N. J., Pelli, D. G., Kurshan, P., and Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vis. Res.* 42, 1165–1184. doi: 10.1016/S0042-6989(02)00045-7
- Marsack, J. D., Thibos, L. N., and Applegate, R. A. (2004). Metrics of optical quality derived from wave aberrations predict visual performance. *J. Vis.* 4, 322–328. doi: 10.1167/4.4.8
- Oruç, I., and Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *J. Vis.* 9, 1–19. doi: 10.1167/9.9.4
- Parish, D. H., and Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise and the efficiency of letter discrimination. *Vis. Res.* 31, 1399–1415. doi: 10.1016/0042-6989(91)90060-1
- Pelli, D. G., and Farrel, B. (1999). Why use noise? *J. Opt. Soc. Am. A* 16, 647–653. doi: 10.1364/JOSAA.16.000647
- Pepose, J. S., and Applegate, R. A. (2005). Making sense out of wavefront sensing. *Am. J. Ophthalmol.* 139, 335–343. doi: 10.1016/j.ajo.2004.11.010
- Porter, J., Guirao, A., Cox, I. G., and Williams, D. R. (2001). Monochromatic aberrations of the human eye in a large population. *J. Opt. Soc. Am. A* 18, 1793–1803. doi: 10.1364/JOSAA.18.001793
- Prins, N., and Kingdom, F. A. A. (2009). *Palamedes: Matlab Routines for Analyzing Psychophysical Data*. Available online at: <http://www.palamedestoolbox.org>
- Ravikumar, S., Bradley, A., and Thibos, L. N. (2010). Phase changes induced by optical aberrations degrade letter and face acuity. *J. Vis.* 10:18. doi: 10.1167/10.14.18
- Rocha, K. M., Vabre, L., Harms, F., Chateau, N., and Krueger, R. R. (2007). Effect of Zernike wavefront aberrations on visual acuity measured using electromagnetic adaptive optics technology. *J. Refract. Surg.* 23, 953–959.
- Rouger, H., Benard, Y., and Legras, R. (2010). Effect of monochromatic induced aberrations on visual performance measured by adaptive optics technology. *J. Refract. Surg.* 26, 578–587. doi: 10.3928/1081597X-20090901-01
- Sawides, L., Gamba, E., Pascual, D., and Marcos, S. (2010). Visual performance with real-life tasks under Adaptive-Optics ocular aberration correction. *J. Vis.* 10, 1–12. doi: 10.1167/10.5.19



- Solomon, J. A., and Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature* 369, 395–397. doi: 10.1038/369395a0
- Thibos, L. N., Hong, X., Bradley, A., and Applegate, R. A. (2004). Accuracy and precision of objective refraction from wavefront aberrations. *J. Vis.* 4, 329–351. doi: 10.1167/4.4.9
- Watson, A. B., and Ahumada, A. J. (2008). Predicting visual acuity from wavefront aberrations. *J. Vis.* 8:17. doi: 10.1167/8.4.17
- Watson, A. B., and Ahumada, A. J. (2012). Modelling acuity for optotypes varying in complexity. *J. Vis.* 12:19. doi: 10.1167/12.10.19
- Young, L. K., Liversedge, S. P., Myers, R. M., Love, G. D., and Smithson, H. E. (2011). Not all aberrations are equal: Reading impairment depends on aberration type and magnitude. *J. Vis.* 11:20. doi: 10.1167/11.13.20
- Young, L. K., Love, G. D., and Smithson, H. E. (2013a). Accounting for the phase, spatial frequency and orientation demands of the task improves metrics based on the visual strehl ratio. *Vis. Res.* 90, 57–67. doi: 10.1016/j.visres.2013.06.007
- Young, L. K., Love, G. D., and Smithson, H. E. (2013b). Different aberrations raise contrast thresholds for single-letter identification in line with their effect on cross-correlation-based confusability. *J. Vis.* 13:12. doi: 10.1167/13.7.12
- Zhao, H.-X., Xu, B., Li, J., Dai, Y., Zhang, Y.-D., and Jiang, W.-H. (2009). Effects of different Zernike terms on optical quality and vision of human eyes. *Chin. Phys. Lett.* 26:054205. doi: 10.1088/0256-307X/26/5/054205

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 March 2014; accepted: 04 September 2014; published online: 30 September 2014.

Citation: Young LK and Smithson HE (2014) Critical band masking reveals the effects of optical distortions on the channel mediating letter identification. *Front. Psychol.* 5:1060. doi: 10.3389/fpsyg.2014.01060

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Young and Smithson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## ADVANTAGES OF PUBLISHING IN FRONTIERS



### FAST PUBLICATION

Average 90 days  
from submission  
to publication



### COLLABORATIVE PEER-REVIEW

Designed to be rigorous –  
yet also collaborative, fair and  
constructive



### RESEARCH NETWORK

Our network  
increases readership  
for your article



### OPEN ACCESS

Articles are free to read,  
for greatest visibility



### TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



### GLOBAL SPREAD

Six million monthly  
page views worldwide



### COPYRIGHT TO AUTHORS

No limit to  
article distribution  
and re-use



### IMPACT METRICS

Advanced metrics  
track your  
article's impact



### SUPPORT

By our Swiss-based  
editorial team