OBJECT DETECTION IN A NOISY SCENE*

Albert J. Ahumada, Jr.
NASA Ames Research Center, Human and Systems Technologies Branch
Moffett Field, California, 94035-1000

Bettina L. Beard
University of California, School of Optometry
Berkeley, California, 94720-2020

Abstract

Observers viewed a simulated airport runway landing scene with an obstructing aircraft on the runway and rated the visibility of the obstructing object in varying levels of white fixed-pattern noise. The effect of the noise was compared with the predictions of single and multiple channel discrimination models. Without a contrast masking correction, both models predict almost no effect of the fixed-pattern noise. A global contrast masking correction improves both models' predictions, but the predictions are best when the masking correction is based only on the noise contrast (does not include the background image contrast).

Keywords: image quality, target detection, noise, vision models, contrast sensitivity

1. Introduction

Object detection typically involves search and pattern recognition in a range of backgrounds. Visual object detection is fundamentally limited by background-induced contrast masking. When the object is present or absent in a constant background, contrast masking can be measured as the discriminability between two images. We are evaluating the ability of image discrimination models to predict object visibility with a fixed background image. If the models are successful, they predict the upper limit of observer performance in an object detection task.

Ahumada, Rohaly, and Watson (SPIE 1995)[1] applied discrimination models to object detection in natural backgrounds. We reported that the detectability of tank targets was better predicted by a multiple channel model than by a single channel model. We then added a simple correction for masking based on visible contrast energy. It improved the predictions for both models and equalized their performance.[2,3,4]

---

*Published in B. Rogowitz and J. Allebach, eds., Human Vision, Visual Processing, and Digital Display VII, SPIE Proceedings Volume 2657, SPIE, Bellingham, WA, Paper 23.

Some object detection situations involve noisy displays. Here we measure object detectability in a complex image masked by fixed-pattern noise. We compare these measurements with discrimination model predictions. Without the masking correction, the single channel model predicts no effect of noise and the multiple channel model predicts masking only by the noise in the channels affected by the object. So, neither model correctly predicts the effect of the fixed-pattern noise. With the masking correction, both models' predictions are improved. The predictions are even better when the masking correction is based only on the noise contrast and does not include the background image contrast.

2. Experiment

2.1 Methods

2.1.1 Stimuli. Two digital images of a simulated airport scene were generated. Image I_1, shown at the top of Figure 1, has an obstructing aircraft on the runway. Image I_0, shown in the middle of Figure 1, is the same image without the obstructing aircraft. We used a single fixed-pattern white noise mask N with uniformly distributed pixel values. Images for the experiment were constructed from these images by adding the background image, a fraction p of the difference between the background and the object images, and a fraction q of the noise image,

$$I_{p,q} = I_0 + p\,(I_1 - I_0) + q\,N + (1-q)\,\bar{N}\,. \tag{1}$$

$\bar{N}$ is the mean of the noise image. A fraction of $\bar{N}$ is added to keep the mean luminance constant. Images were generated for the six p values 0, 0.05, 0.10, 0.20, 0.40, and 1, and for the q values 0, 0.25, 0.50, and 1.0. The image at the bottom of Figure 1 illustrates the case of $p = 1$ and $q = 0.5$. The 128×128 pixel gray-scale images were presented on a 15 inch Sony color monitor whose luminance in cd/m^2 was closely approximated by

$$L = 0.05 + (\,0.024\,d\,)^{2.4}\,, \tag{2}$$

where d is the digital image pixel value. The mean luminance of the images and surrounding screen region was about 10 cd/m^2. The viewing distance of 127.5 cm and the image size of 6.0 cm give a viewing resolution of 47.5 pixels per degree of visual angle. The plane/runway scene thus subtended 2.7 deg visual angle, the plane alone fit in a rectangle 0.78 deg by 0.17 deg of visual angle (37 horizontal and 8 vertical pixels). It affected a total of 96 pixels. When an image was not present, the screen was filled with random, uniformly distributed, gray scale pixels. Because the display had only 32 different levels of gray scale (IBM-PC compatible VGA display mode) the no-noise condition was run at twice the digital image contrast to allow more dynamic range. The image duration was 1.0 second.

2.1.2 Observers. Four female observers, aged 18 to 37 years, with corrected acuity of 20/20 or better were tested.

2.1.3 Procedure. The observers were asked to rate each image on a 4 point rating scale according to the following interpretation:

1-Definitely did not have a plane.
2-Probably did not have a plane.
3-Probably did have a plane.

4-Definitely did have a plane. In addition, the observers were asked to try to use the 4 response categories with roughly equal frequency.

Within a block of 60 trials, the mask noise level q was held constant, while the four object/background p levels occurred randomly (with probability 0.25). Table 1 shows the four values of p used at each q value (the coefficient determining the noise level).

Table 1 - Signal level values p used at each noise level value q				
q	p's			
0	0	0.05	0.1	0.2
0.25	0	0.05	0.1	0.2
0.5	0	0.1	0.2	0.4
1	0	0.2	0.4	1.0

Groups of four repetitions of the four noise levels were independently sequenced using 4×4 Latin squares. Observers 1 and 2 completed 16 repetitions of each noise level, Observer 3 completed 8 repetitions, and Observer 4 completed 10 repetitions in 5×5 Latin squares, including a no-noise condition at the same contrast as the noise conditions.

2.2 Data analysis

2.2.1 Method. For a given noise level, the distance d' in discriminability units from each object image to its non-object image was measured in the context of a one-dimensional Thurstone scaling model.[5] The scaling model has the following assumptions:

1. The presentation of an image generates an internal value that is a sample from a normal distribution with unit variance.

2a. The mean of the distribution generated by a background image I_0 is zero.
2b. The mean of the distribution generated by an original object image I_1 is d'.
2c. The mean of the distribution generated by an image I_p is $p\,d'$.

3. The observer has 3 fixed criteria that are used to categorize an internal value to one of the 4 responses.

The scaling model for this experiment has 4 d' parameters and 3 category boundaries for each observer. Parameters were estimated by the method of maximum likelihood separately for each block.

2.2.2 Experimental results. Median d' estimates for each observer and for the 4 noise levels are given in Table 2.

Table 2 - Median experimental discriminability indices d'				
noise level q	0	0.25	0.5	1
Observer 1	18.5	11.9	6.6	3.3
Observer 2	24.9	11.6	8.8	4.1
Observer 3	24.4	9.5	8.8	5.8
Observer 4	28.4	15.4	9.0	5.2
Geometric mean	24.8	11.9	8.2	4.5

The standard deviation of an individual score in decibels (dB = 20 × the log of the score) is estimated to be 1.3 dB, based on the observer by noise level interaction, which has 9 degrees of freedom. This leads to 95% confidence intervals of ±1.4 dB for the means for each noise level. Figure 2 plots the data of Table 2 with the confidence intervals about the means. Observer 4 had a median d' of 18.4 for the no-noise condition at the same contrast as the noise conditions, only slightly higher than her d' value of 15.4 for the $q = 0.25$ condition. The large difference from the $q = 0$ and the $q = 0.25$ conditions is seen to be mainly an effect of the lower signal level in the noise conditions.

3. Models

3.1 Algorithms

3.1.1 Multiple channel model. The multiple channel model is based on the Cortex transform of Watson.[6] It is similar in spirit to his original multiple channel model,[7] and is similar in detail to others based on the Cortex transform.[8,9,10]

The multiple channel model calculation for a pair of images (I_0 and I_1) has the following steps. The images I_1 and I_0 are converted to luminance images by the calibration function of Equation (2). The images are converted to luminance contrast by subtracting and then dividing by the background image mean luminance $\bar{L}_0$,

$$I_j \leftarrow (I_j - \bar{L}_0) / \bar{L}_0 . \tag{3}$$

The operations on the image indicate the operation applied separately to each pixel. A contrast sensitivity function (CSF) filter S is then applied to the two contrast images.

$$I_j \leftarrow F^{-1}[S\, F[I_j]] , \tag{4}$$

where F and F^{-1} are the forward and inverse Fourier transforms. Next the Cortex transform is applied to the images resulting in coefficients $C_{j,k}$, where the index k ranges over spatial frequency, orientation, and spatial location. The detectability d_k contributed by the kth spatial frequency, orientation, and position is then computed as the absolute value of the difference in the Cortex transform coefficients, masked by the background coefficient if it is above threshold.

$$\begin{aligned} d_k &= | C_{1,k} - C_{0,k} | , \quad \text{if } C_{0,k} \le 1.0 , \\ d_k &= | C_{1,k} - C_{0,k} | \;/\; | C_{0,k} |^{0.7} , \text{ if } C_{0,k} > 1.0 . \end{aligned} \tag{5}$$

Finally, d' is given by a Minkowski sum of the individual contributions with summation exponent β,

$$d' = \left(\sum_k d_k^{\beta}\right)^{1/\beta}. \tag{6}$$

For the case that $\beta = \infty$, the result is the largest of the d_k.

3.1.2 Single channel model. For the single channel model, the steps are the same through the image filtering, then the filtered image values are used to compute

$$d_k = |I_{1,k} - I_{0,k}|, \tag{7}$$

where the index k now refers to image pixels. Equation (6) is then used to obtain d'.

3.1.3 Contrast normalization. Without the correction factor, the single channel model predicts no contrast masking at all and the multiple channel model only predicts masking within the channels affected by the signal. Recent work demonstrates masking by contrast energy in channels not containing the signal.[11] New versions of the multiple channel models incorporating lateral interactions among cortical unit channels to account for between-channel masking have been developed.[12–15] A model similar to theirs would result by replacing Equation (5) with

$$d_k = |C_{1,k} - C_{0,k}| \frac{c_0}{\left(c_0^{a_0} + \sum_{k'} c_{k,k'} |C_{0,k'}|^{a_{k,k'}}\right)^{1/a_0}}, \tag{8}$$

where c_0 and a_0 are constants, $c_{k,k'}$ represents the weight of the masking of channel k' on channel k, and $a_{k,k'}$ represents the growth of that masking with the activity in channel k'. If we make the simplifying assumptions that the $c_{k,k'}$ are all equal and sum to unity, that the $a_{k,k'} = 2$, and $a_0 = 2$, the result is that the factor multiplying the difference term is no longer a function of k and can be factored out of the Minkowski metric Equation (6). Also, the Cortex transform has the property that the sum of squares of the coefficients equals the sum of squares of the image values, so the simplification assumptions result in the d' prediction formula,

$$d' = d'_{unmasked} \frac{c_0}{\sqrt{c_0^2 + c^2}}, \tag{9}$$

where $d'_{unmasked}$ is computed from the unmasked differences, c is the RMS background image contrast passed by the CSF filter, and c_0 is a parameter representing the contrast level at which the masking becomes effective. To compute c, the CSF is normalized to unity at its peak value. Instead of dealing with the additional computational complexity and parameter estimation problems of Equation (8), we will simply use Equation (9) to correct the predictions of the single and multiple channel models.

3.2 Model parameters

The model parameters used are those that proved to be best in previous studies.[1–4] The CSF filters were calibrated to agree with the CSF formula developed by Barten.[16] The filters have a difference of Gaussian form,

$$S(f) = a_c \exp^{-(f/f_c)^2} - a_s \exp^{-(f/f_s)^2}, \tag{10}$$

where a_c and a_s are the center and surround amplitude parameters and f_c and f_s are the center and surround frequency cutoff parameters. Table 3 gives the CSF and β parameters for the multiple channel and the single channel models. The amplitude parameters have the dimensions of JND's per unit contrast and the cutoff parameters have the dimensions of cycles per degree of visual angle.

Table 3 - Model parameters

channels	β	a_c	f_c	a_s/a_c	f_c/f_s
multiple	4	15.5	20.8	0.77	5.6
single	4	18.5	16.4	0.68	7.9

3.3 Model predictions and results

3.3.1 Predictions without a contrast masking correction. The model predictions for d' without a contrast masking correction given in Table 4 for each of the four noise levels.

Table 4 - Model d''s without a contrast masking correction

noise level q	0	0.25	0.5	1
multiple channel	4.0	2.3	2.2	1.9
single channel	24.5	11.5	11.5	11.8

Figure 3 shows the predictions of Table 4 plotted with the mean observer results. Both models correctly predict the difference between $q = 0$ and $q = 0.25$ caused by scaling the down the aircraft image to make room for the noise. The single channel model predicts no masking by the noise. The multiple channel model predicts very little masking by the noise. Table 5 shows the sensitivity scale factors needed to equalize the average log predictions of the models and the observers. It also shows the average error of prediction in decibels using the scale factor and an F statistic representing the statistical goodness-of-fit of the error. The multiple channel model averages a factor of 4 too insensitive, while the single channel average sensitivity is within the range of that of the observers. The underprediction of the masking effects causes the errors to be large. Both F's are highly significant, since the 99.9 percentile of the F distribution with 3 and 9 degrees of freedom is 13.9.

Table 5 - Model fits without contrast masking correction

model	scale factor	error, dB	F
multiple channel	4.1	3.5	30.5
single channel	0.72	4.0	38.5

3.3.2 With contrast masking correction. RMS contrast values for normalizing the d' values are shown in Table 6 for each of the 4 noise plus background images, filtered by the CSF for each model.

Table 6 - RMS image contrast				
noise level q	0	0.25	0.5	1
multiple channel	0.136	0.076	0.098	0.158
single channel	0.150	0.079	0.093	0.136

Figure 4 shows the predictions of Figure 3 corrected with a c_0 of 0.04 and the RMS contrast values of Table 6. Now both models predict the effect of the noise better when the noise is present, but they predict too much masking of the target by the image alone. Table 7 shows the goodness-of-fit measures as in Table 5. The scale factors show that now both models predict too much masking.

Table 7 - Model fits with contrast masking correction			
model	scale factor	error, dB	F
multiple channel	12.2	3.3	25.3
single channel	2.1	3.8	34.7

3.3.3 Contrast masking correction based on noise alone. The poor fit above is what one might expect from using an image-wide estimate for image masking while the runway region has little contrast variation. The values of Table 6 can be decomposed to show that the RMS visible contrast from the full ($q = 1$) noise alone is 0.144 for the multiple channel model and 0.114 for the single channel model. Figure 5 shows the predictions of Figure 3 corrected with a c_0 of 0.04 and the noise component of the RMS visible contrast. Now both models fit well, with a slight error in the direction that would result from a small image masking effect. Table 8 shows the goodness-of-fit measures as in Table 5. Now both models have scale factors close to unity and the single channel model fits the noise effect quite well. The multiple channel F now barely exceeds the 99th percentile of the F distribution (6.99), and the single channel F is just above the 90th percentile (2.81).

Table 8 - Model fits using only noise in the contrast masking correction			
model	scale factor	error, dB	F
multiple channel	1.30	1.7	7.02
single channel	1.16	1.1	2.84

4. Discussion

The improvement in the model predictions resulting from limiting the contrast masking correction to the noise, suggests that the contrast masking correction should be based on the contrast in a smaller region containing the target object. We had success before[2–4] with the correction based on the same sized image, and experiments measuring contrast effects on perceived contrast indicate considerable spatial spread.[19–22] Current models[12–15] extend the masking interactions only to channels differing in orientation at the same location and spatial frequency. Also recent attempts to measure contrast masking by a surround masker found none.[23,24] We currently recommend that the contrast masking correction be based on an estimate of the image contrast in the immediate region of the target object.

The results demonstrate that the single channel model with an appropriate contrast masking correction can outperform the multiple channel model with or without a general gain control. Although a multiple channel model with inter-channel interactions might do better in this situation, it probably would require more strongly oriented signals and maskers to obtain a benefit for the extra calculations. One problem with the contrast masking correction and the multiple channel model is that contrast in the signal channels contributes to masking twice. The multiple channel model might be the better of the two with the correction if, for example, the within-channel masking exponent and the correction exponent were both lowered. The results here show that even though the single channel model does not predict the details of oriented contrast masking, such as the results of Foley,[11] it can be a useful alternative to more complicated models.

5. Acknowledgments

Ren-Sheng Horng wrote the experimental display and response collection program. Andrew Watson wrote the basic Mathematica routines that generated the model and metric predictions and made helpful suggestions. We are also grateful for the help of Ann Marie Rohaly, Cynthia Null, Jeffrey Mulligan, and Robert Eriksson. This work was supported in part by NASA Grant 199-06-39 to Andrew Watson and NASA Aeronautics RTOP #505-64-53.

6. References

1. A.J. Ahumada, Jr., A.B. Watson, A.M. Rohaly (1995) Models of human image discrimination predict object detection in natural backgrounds, in B. Rogowitz and J. Allebach, eds., *Human Vision, Visual Processing, and Digital Display IV*, Proc. Vol. 2411, SPIE, Bellingham, WA, pp. 355-362.
2. A.J. Ahumada, Jr., A.M. Rohaly, A.B. Watson (1995) Image discrimination models predict object detection in natural backgrounds, *Investigative Ophthalmology and Visual Science,* vol. 36 (ARVO Suppl.), p. S439 (abstract).
3. A.M. Rohaly, A.J. Ahumada, Jr., and A.B. Watson (1995) A Comparison of Image Quality Models and Metrics Predicting Object Detection, *SID Digest*, *26*, 45-48.
4. A.J. Ahumada, Jr., A.B. Watson, A.M. Rohaly (1995) Object detection in natural backgrounds predicted by discrimination performance and models *Perception*, Vol. 24, ECVP Suppl., p. 7 (abstract).
5. W.S. Torgerson (1958) *Theory and Methods of Scaling*, Wiley, New York.
6. A.B. Watson (1987) The Cortex transform: rapid computation of simulated neural images, *Computer Vision, Graphics, and Image Processing, 39*, 311-327.
7. A.B. Watson (1983) Detection and recognition of simple spatial forms, in O. J. Braddick and A. C. Sleigh, eds., *Physical and biological processing of images,* Springer-Verlag, Berlin.
8. A.B. Watson (1987) Efficiency of an image code based on human vision. *JOSA A, 4*, 2401-2417.
9. S. Daly (1993) The visible differences predictor: an algorithm for the assessment of image

fidelity, in Watson, ed. *Digital Images and Human Vision*. MIT Press, Cambridge, MA.
10. J. Lubin (1993) The use of psychophysical data and models in the analysis of display system performance, in Watson, ed. *Digital Images and Human Vision*. MIT Press, Cambridge, MA.
11. J.M. Foley (1994) Human luminance pattern-vision mechanisms: masking experiments require a new model, *Journal of the Optical Society of America A*, vol. 11, pp. 1710-1719.
12. P.C. Teo, D.J. Heeger (1994) Perceptual image distortion, in B. Rogowitz and J. Allebach, eds., *Human Vision, Visual Processing, and Digital Display V*, Proceedings Volume 2179, SPIE, Bellingham, WA, pp. 127-141.
13. P.C. Teo, D.J. Heeger (1994) Perceptual image distortion, *Proceedings of ICIP-94*, Volume II, IEEE Computer Society Press, Los Alamitos, California, pp. 982-986.
14. P.C. Teo, D.J. Heeger (1995) A general mechanistic model of spatial pattern detection, *Investigative Ophthalmology and Visual Science*, vol. 36, no. 4 (ARVO Suppl.), p. S438 (abstract).
15. A.B. Watson, J.A. Solomon (1995) Contrast gain control model fits masking data, *Investigative Ophthalmology and Visual Science*, vol. 36, no. 4 (ARVO Suppl.), p. S438 (abstract).
16. P.G.J. Barten (1993) Spatiotemporal model for the contrast sensitivity of the human eye and its temporal aspects, in B. Rogowitz and J. Allebach, eds., *Human Vision, Visual Processing, and Digital Display IV*, Proc. Vol. 1913, SPIE, Bellingham, WA, pp. 2-14.
19. M.W. Cannon, S.C. Fullenkamp (1991) Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations, *Vision Research*, vol. 31, pp. 1985-1998.
20. J.S. DeBonet, Q. Zaidi (1994) Weighted spatial integration of induced contrast-contrast, *Investigative Ophthalmology and Visual Science*, vol. 35 (ARVO Suppl.), p. 1667.
21. B. Singer, M. D'Zmura (1994) Color contrast induction, *Vision Research*, vol. 34, pp. 3111-3126. 22. M. D'Zmura, B. Singer, L. Dinh, J. Kim, J. Lewis (1994) Spatial sensitivity of contrast induction mechanisms, *Optics and Photonics News*, vol. 5, no. 8 (suppl), p. 48 (abstract).
23. R.J. Snowden, S.T. Hammett (1995) The effect of contrast surrounds on contrast centres, *Investigative Ophthalmology and Visual Science*, vol. 36, no. 4 (ARVO Suppl.), p. S438 (abstract).
24. J.A. Solomon, A.B. Watson (1995) Spatial and spatial frequency spreads of masking: Measurements and a contrast-gain-control model, *Perception*, Vol. 24, ECVP Suppl., p. 37 (abstract).

Figure 1. (Top) Airport scene with an obstacle aircraft on the runway.
(Middle) The same scene without the aircraft.
(Bottom) The aircraft scene (p=1) masked by the noise at q=0.5.

Figure 2. Object detectability data from 4 observers for 4 noise levels.

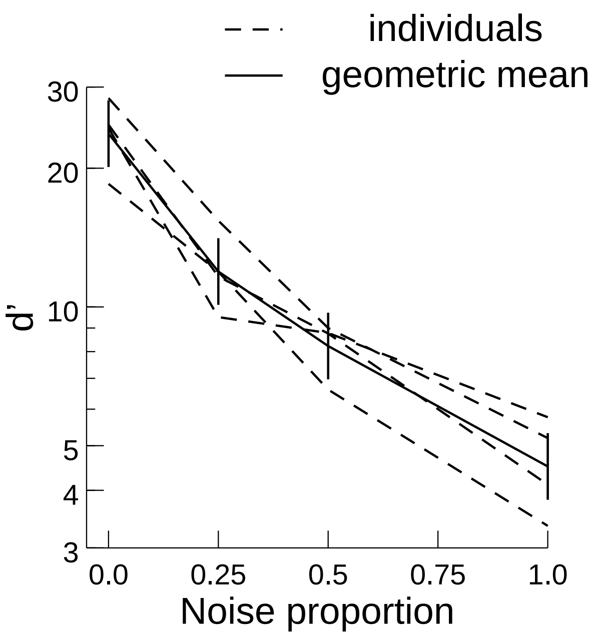

Figure 3. Predictions of scaled models without contrast masking correction.

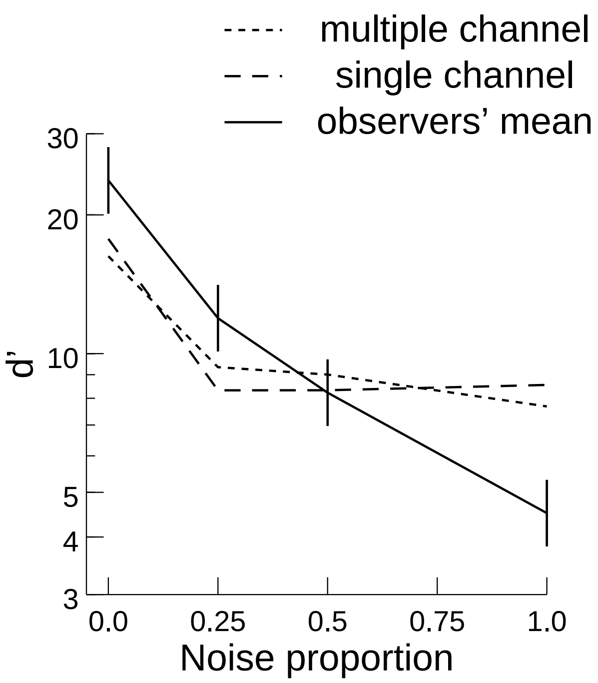

Figure 4. Predictions of scaled models with the contrast masking correction.

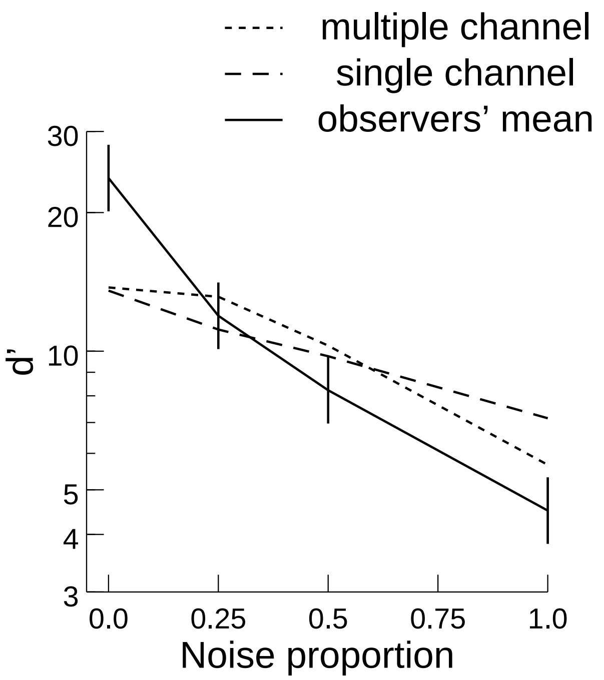

Figure 5. Predictions of scaled models with the contrast masking correction based only on the noise.