Transducer Function for Energy (Viewprint) Models

Stanley A. Klein* and Thom Carney*#
*School of Optometry, UC Berkeley, #Neurometrics Institute, Berkeley, CA

Abstract. Energy based, joint space-spatial frequency representations are useful for image fidelity metrics. Elimination of local phase (in energy models) produces an accelerated transducer function that is steeper than quadratic at low contrast and is linear at high contrast. This is a simple way to get a realistic transducer function.

Background of medium bandwidth mechanisms and the space-spatial frequency representation. In the late 1960's Fergus Campbell and colleagues developed an approach to vision modeling that was based on the spatial frequency content of the image The experimental basis of this approach was the finding that a square wave grating above about 4 c/deg was detectable when its fundamental component was detectable (Campbell & Robson, 1968). It appeared that the visibility of test patterns was based on the visibility of each Fourier component separately. In an effort to push this approach further, Sachs, Nachmias & Robson (1971) measured the visibility of grating pairs that were close in spatial frequency. They found that even at the closest frequencies (frequency ratios of 1.25) the gratings did not summate well, indicating the underlying mechanisms were very narrowly tuned. However, Stromeyer & Klein (1974) suggested that the lack of summation of closely spaced components could be due to spatial probability summation across the beats such that the visibility is based not just on the peak, but also on the lower contrast portions of the pattern. This conclusion was verified by Stromeyer & Klein (1975) who used frequency modulated patterns rather than the amplitude modulated patterns used by Sachs et al. (1971). The conclusion of our studies was that the underlying mechanisms were not very narrow. They had medium bandwidths (full bandwidths around 1.5 octaves). We go through this history for two reasons: (1) In describing their frequency modulated gratings, Stromeyer & Klein (1975) introduced the concept of space-spatial frequency plots, which is the theme of the present paper. We believe that it is often useful to view a pattern in its space-spatial frequency representation. (2) The space-spatial frequency representation depends on the assumed bandwidth of the underlying mechanisms. The above discussion shows why we assume the underlying mechanisms have medium bandwidth.

Uses and validation of the viewprint representation. The present paper is concerned with a variant of the space-spatial frequency representation in which absolute phase information is eliminated. One might worry that eliminating phase is in conflict with available data on position thresholds. Phase is closely coupled to position. It might, for example, be thought that our ability to do position tasks is based on comparing the relative stimulation of even and odd symmetric mechanisms. Klein and Levi (1985) took the opposite approach. We developed an "energy" model for spatial vision. In this model the detection stage only saw the Pythagorean sum of the even and odd mechanism which erases information about local phase. In the language of signal detection theory, detection would be based not on the "stimulus known exactly" ideal observer, but rather on the "stimulus known except for phase" ideal observer. We worried that this latter model would not perform well on position (hyperacuity) tasks so we carried out a number of bisection experiments. Under some conditions we found position thresholds of less than 1 sec of arc (these thresholds are so small that they now hold the Guinness record). Our energy model was fully able to predict thresholds this small (see Klein, 19923a and 1993b for further details on predicting the 1 sec of arc value).

Figure 1

We called our energy model the "viewprint" model because of its similarity to the "voiceprint" models of audition. The most familiar example of a voiceprint is shown in Fig. 1. On the horizontal axis is time and on the vertical axis is temporal frequency. The representation shown in Fig. 1 erases phase information. The frequency, but not the phase, of each note is specified. This is an excellent representation for music for both the musician (it would be difficult for the musician to produce chords with phase specified exactly) and for the listener (the human auditory system can not discriminate non-harmonically related phases). If four frequencies are simultaneously presented, the relative phases are irrelevant to the listener. A musician looking at Fig. 1 is able to imagine the sound. That would not be possible by looking at the oscilloscope trace of the music.

It is our belief that spatial vision works the same way as audition. For closely spaced objects (less than about 5 min apart, which is about the size of a cortical hypercolumn) we believe a spatial frequency (*relative* position or size) representation is used, and for features farther apart we believe an *absolute* spatial representation is used. So far what we are saying sounds just like the DCT transform that is used in JPEG and MPEG. But whereas the DCT transform maintains the phase of the stimulus we take the Pythagorean sum of even and odd symmetric mechanisms at each point. The DCT (or wavelet) representations are ideal for image reconstruction following compression because they are invertable. However, these representations may not be ideal for image quality assessment because a slight shift of the image can produce dramatic changes in the DCT coefficients and thus make large contributions to the error metric, while to a human observer the shifted image is perceptually unchanged. Consider, for example, a 10 c/deg grating that is shifted by 180 deg. The viewprint representation, in agreement with human perception and in disagreement with the DCT or wavelet representation, would not detect the shift. The viewprint representation is ideal for keeping excellent relative position information (the relative activity of mechanisms tuned to different sizes) while eliminating absolute position (the Pythagorean sum eliminates local phase information). It is for their usefulness in image quality metrics that we examine energy models that ignore phase.

Fig. 2 shows the space-size layout with even and odd pairs at each point. Shown in Fig. 2 are Cauchy function pairs that have identical spatial frequency tuning, which facilitates taking the Pythagorean sum (Klein & Levi, 1985).

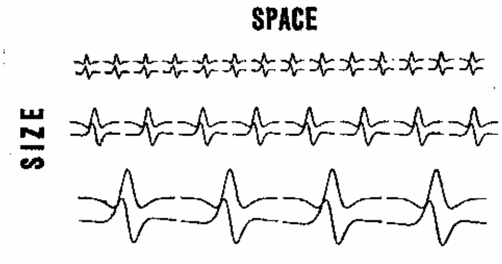

Figure 2

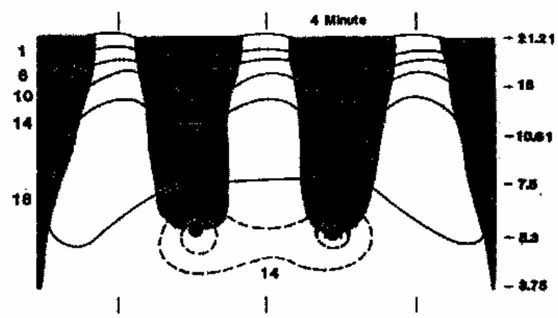

Figure 3

Fig. 3 shows a viewprint for one of our bisection three-line (Klein & Levi, 1985). The abscissa is spatial position. The tic marks at the top and bottom of the figure indicate the location of the three white lines that are spaced by 4 min. The shading indicates where the even symmetric mechanisms are stimulated a negative amount (in the regions between the stimulus lines). The numbers on the right side of the figure indicate the peak spatial frequencies of the underlying mechanisms at that vertical location. The contour lines are the iso-d' lines of a "made up" transducer function. The d' levels are indicated on the left side of the plot. The transducer function that was used had the form:

$$d' = 2.5 \ln(1 + .5\, s^2) \qquad (1)$$

where s is the mechanism stimulation strength in threshold units. This transducer function was ad hoc. It was chosen to give a d'=1 for s=1 (this would be threshold if thresholds were based just on the peak responding mechanism), a 10% Weber fraction at high values of s ($\Delta s/s \approx .1$), and a quadratic dependence on s for small values of s. One purpose of the present paper is to probe deeper into the nature of this phase uncertainty task. We ask what the ideal "energy" (stimulus known except for phase) observer's transducer function is like, rather than the empirically based function in Eq. 1.

Green & Swets (1966) were interested in exactly the same question that we are asking, except their interests were in the domain of audition rather than vision. As was discussed in connection with Fig. 1 human observers are poor at detecting the phase of a stimulus. Thus Green and Swets were interested in ideal observers that ignore phase and base their judgments on the amplitude or envelope (the energy) of the waveform. Consider the task of detecting a sinusoid in noise. Fig. 4 presents a graphical representation of the problem.

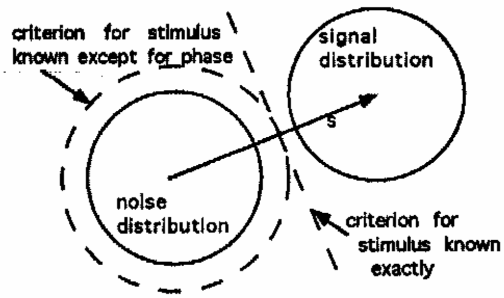

Figure 4

The x and y axes represent the response of symmetric and antisymmetric mechanisms respectively. The solid circle centered at the origin represents a contour plot of the mechanism's activity when a signal is not present. The displaced circle represents the mechanisms' activity in the presence of a signal. The displacement direction is determined by the phase of the signal and the displacement magnitude is determined by the signal energy. The dotted straight line is the criterion location that would be used by a "stimulus known exactly" ideal observer. The dotted circle is the criterion location that would be used by a "stimulus known except for phase" ideal observer (Geisler & Davila 1985 also discuss the "stimulus know statistically"). Fig. 4 shows that the criterion placement in the latter case is not very efficient, especially at low stimulus strengths. To quantify the relative efficiency of the two tasks we carried out a Monte Carlo simulation using Matlab (Mathworks). In order to not have our conclusions depend on the precise position of the criteria, we simulated a two-alternative forced choice.

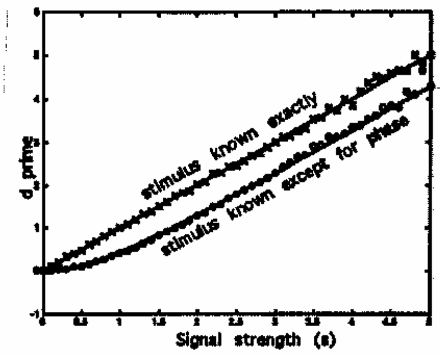

Figure 5

Fig. 5 shows the transducer function (d' vs. stimulus strength) for the two types of ideal observer (stimulus known exactly and stimulus known except for phase). The program producing Fig. 5 is presented in the following Matlab code:

```
1  runs=10000;
2  for istreng = 1:51; s(istreng)=(istreng-1)/10;
3    x=randn(4, runs);
4    x(4,:)=x(4,:)+ s(istreng)*ones(1,runs);
5    prob(1,istreng)=sum(x(4,:)>x(2,:))/runs;
6    Pn=(x(1,:).^2+x(2,:).^2);
7    Ps=(x(3,:).^2+x(4,:).^2);
8    prob(2,istreng)=sum(Pn < Ps)/runs;
9  end
10 d=erfinv(prob*2-1)*2;
11 plot(s,d(1,:),'x',s,d(2,:),'o');hold on
12 d(1,:)=s.^3 ./(s.^2 + .5*s + .8*sqrt(s));
13 d(2,:)=s; plot(s,d,'-')
```

Line 1: Specifies 10,000 two-alternative forced choice trials.

Line 2: The stimulus strength goes from 0 to 5 in steps of .1 (there are 51 strengths).

Line 3: Four Gaussian random numbers (mean=0 and sd=1) are chosen for each trial. Numbers 1 and 2 are for the odd and even channels for the noise trial, numbers 3 and 4 are for the odd and even channels for the signal trial.

Line 4: The signal is added to the even channel (the fourth random number). For this case the signal distribution in Fig. 4 would be on the x-axis.

Line 5: The probability correct for the signal known exactly. This is given by the percentage of times that a random number in the fourth row is greater than one in the second row.

Lines 6 and 7: The energy (Pythagorean sum of even and odd channels) for the noise and the signal presentations respectively.

Line 8: The probability correct for the signal known except for phase. Note the similarity to line 5.

Line 9: End of the loop.

Line 10: Converts probability correct to z-scores which are then multiplied by √2 to get d' for a 2 alternative forced choice task. The factors of 2 are needed to convert the Matlab error function to z-scores.

Line 11: Plots the data.

Line 12: An approximate formula that does a decent job of fitting the energy model transducer.

Line 13: Plots the fitting functions.

The solid line shows that for the stimulus known exactly d' is equal to the signal strength. For the stimulus known except for phase, the d' transducer function is well fit by:

$$d'=s^3/(s^2 + 0.5\ s + 0.8\ s^{1/2}) \qquad (2)$$

The values of 0.5 and 0.8 were arrived at by trial and error. Three aspects of Eq. 2 are unusual:

1) To our knowledge this is the first time the ideal observer energy transducer has been calculated.

2) The factor of 0.7 is connected with the amount of reduction of d' for large stimulus values.

3) At low stimulus strength the transducer function accelerates with a power function of about 2.5. A quadratic acceleration does not fit the data well.

Discussion

We have conjectured that a space-size energy plot is a useful representation for an image fidelity metric.

Space-size representation. In order to produce such a representation one must use mechanisms that are localized both in space and in size (spatial frequency). Wilson & Stork (1990) and Klein & Beutter (1992) discuss issues related to what is required for mechanisms to have joint localization in both space and spatial frequency. The need for localization in space is obvious since image fidelity must be maintained at each spatial point. The need for localization in spatial frequency is less obvious and is based on the finding that masking and detection are spatial frequency and orientation specific.

Energy representation. In order to produce an energy representation one must remove the local phase information by taking the Pythagorean sum across even and odd mechanisms at the same spatial position and with the same spatial frequency tuning. We desire an energy representation in order to allow the fidelity metric to become somewhat resistant to slight distortions of position.

Transducer function for energy. One consequence of an energy representation is that local phase is unknown. At low stimulus strengths the phase uncertainty produces a surprisingly large loss of d'. Our simulations show that the linear transducer function is converted to an accelerating function with an exponent of about 2.5 at low contrasts. This loss of visibility is greater than what would be expected from an ideal observer with two independent channels of uncertainty. The phase unknown case has more than two degrees of uncertainty because the polarity is unknown and also both the symmetric and antisymmetric channels can be activated at once. We found an analytic expression that provides a good match to the transducer function.

References

Campbell, FW & Robson, JG (1968). Application of Fourier analysis to the visibility of gratings. *J. Physiology (London)* **197**, 551-566.

Geisler, WS & Davila, KD (1985). Ideal discriminators in spatial vision: two-point stimuli. *J.Opt. Soc. Am. A.* **2**, 1483-1497.

Green, DM & Swets, JA (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons.

Guinness Book of World Records. (1991). Ed. D. McFarlan. Bantam Books, New York.

Klein, SA & Levi, DM (1985). Hyperacuity thresholds of one second: theoretical predictions and empirical validation. *J.Opt. Soc. Am. A* **2**, 1170-1190.

Klein, S. A. & Beutter, B. (1992). Minimizing and maximizing the joint space-spatial frequency uncertainty of Gabor-like functions. *J. Opt. Soc. Am. A.* **9**, 337-340.

Klein, SA (1993a). Spatial vision models: Problems and successes. Chapter in *Spatial Vision in Humans and Robots*, Cambridge University Press, 9-32.

Klein, SA (1993b). Fidelity metrics and the test-pedestal approach to spatial vision. *Comp-utational Vision Based on Neurobiology*, Teri Lawton, Ed., Proc. SPIE **2054**, 142 - 154.

Sachs, MB, Nachmias, J & Robson, JG (1971). Spatial frequency channels in human vision. *J.Opt. Soc. Am.* **61**, 1176-1186.

Stromeyer, CF & Klein, SA (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Res.* **14**, 1409-1420.

Stromeyer, CF & Klein, SA (1975). Evidence against narrow-band spatial frequency channels in vision: the detectability of frequency modulated gratings. *Vision Res.* **15**, 899-910.

Wilson, HR & Stork, DG (1990). Do Gabor functiuons provide appropriate descriptions of visual cortical receptive fields? *J.Opt. Soc. Am. A* **7**, 1362-1373.

Acknowledgement. This work was supported by NEI grant R01 04776 and AFOSR F49620-95-C-0018.