Forty-four years of studying light adaptation using the probed-sinewave paradigm

S. Sabina Wolfson   Columbia University, Department of Psychology, New York, NY, USA

Norma Graham   Columbia University, Department of Psychology, New York, NY, USA

Here we examine results from 44 years of probed-sinewave experiments investigating the dynamics of light adaptation. We also briefly examine four models that have been tested against the results. In these experiments, detection threshold is measured for a test stimulus superimposed at various times (phases) on a sinusoidally flickering homogeneous background. The results can be plotted as probe-threshold versus phase curves. Overall, the curves from different laboratories are remarkably similar given the substantial differences in experimental parameters. However, at medium frequencies of background flicker, there are some differences between the majority of the studies and a minority of two. An examination of the full set of results suggests that the differences are not as significant as they first appear and that the experimental condition leading to the differences is the use of long wavelength light in the two minority studies. Of the four models that have been tested, two fail to predict important features of the results, another is critically dependent on a mechanism unlikely to exist in the appropriate physiology, and the last seems quite promising.

Keywords: probed-sinewave, light adaptation, computational models, temporal dynamics, flicker

Introduction

Over the past 44 years, numerous researchers have used the *probed-sinewave paradigm* to explore the dynamics of light adaptation, and it has proven powerful for testing models of light adaptation (Boynton, Sturr, & Ikeda, 1961; DeMarco, Hughes, & Purkiss, 2000; Hood & Graham, 1998; Hood, Graham, von Wiegand, & Chase, 1997; Maruyama & Takahashi, 1977; Shady, 2000; Sherman & Spitzer, 2000; Shickman, 1970; Snippe, Poot, & van Hateren, 2000, 2004; Wilson, 1997; Wolfson & Graham, 2000, 2001a, 2001b; Wu, Burns, Elsner, Eskew, & He, 1997).

In the probed-sinewave paradigm, a brief spot of light (called a *probe*) is presented on a sinusoidally flickering (spatially uniform) field of light (called a *background*). Detection threshold for the probe is measured at various times (*phases*) with respect to the flickering background. While the particular stimulus parameters vary across experiments, a typical spatial layout is shown in Figure 1. In this spatial layout, there is a small blurry-edged probe superimposed on a larger background. A typical temporal profile is shown in Figure 2. In this temporal profile, the sinusoid is the luminance profile of the flickering background, and the eight short vertical lines show eight phases at which an incremental probe might be presented (although on a given trial, only one of the probes would be presented).

This article was originally motivated by our concern about certain differences between results collected with two laboratory setups: an optical system with LED light sources (see description in Hood et al., 1997, and the "H" column in Table 1) and a computer-controlled CRT display (see description in Wolfson & Graham, 2000, 2001a, and the "W" column in Table 1). The results from these two setups showed some systematic differences even in conditions we thought were substantially the same for the purposes of this paradigm. Because all the observers showed similar results *within* a particular setup, individual differences among the observers seemed unlikely to explain the differences between the results from the two setups. At about the same time, Snippe et al. (2000) also noticed differences between their results and those of Hood et al. (1997).

Results from the probed-sinewave paradigm discriminated among (and rejected) alternative theoretical models of light adaptation. However, the results from the paradigm also showed unexplained discrepancies between experimental results from different laboratories. Thus, it seemed worthwhile to look carefully at all the published results in the literature. We hoped to discover what general similarities and differences there are and how these similarities and differences relate to the predictions of the several models.

In this article, we present and compare data from Boynton et al. (1961), DeMarco et al. (2000), Hood et al. (1997), Maruyama and Takahashi (1977), Shady (2000), Shickman (1970), Snippe et al. (2000, 2004), Wolfson and Graham (2001a), and Wu et al. (1997). The stimulus and experiment parameters for these data sets are listed in Table 1. These represent all the published studies known to us that use the straightforward version of the probed-sinewave
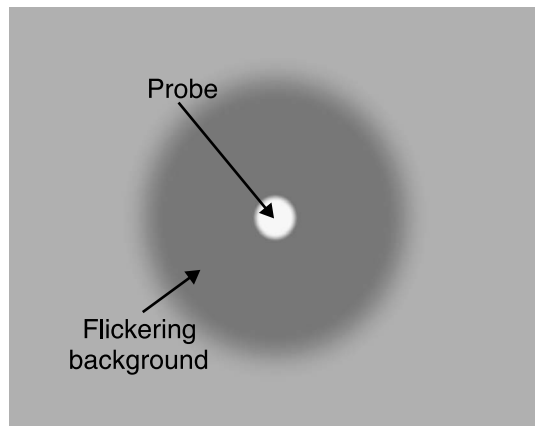
Figure 1. An example of a spatial layout used in the probed-sinewave paradigm, showing a blurry-edged increment probe at a phase of 270 deg.

paradigm with the following exception: When a group has published several studies and the results appear the same in all important respects, we show only one typical data set. (See Appendix for details.)

We find that the discrepancies between the results from different setups are *not* as significant as they first appear. The discrepancies turn out to be more quantitative than qualitative. We suggest that the use of different wavelengths of light by different studies is a large factor in the discrepancies. Lastly, we conclude that the Snippe et al. (2000, 2004) model is currently the most promising of the models applied to the probed-sinewave paradigm.

Methods

The results here are all from previously published papers. The data sets are listed in Table 1, along with various parameters. Some of the data were obtained directly from the authors in electronic format; the rest were manually extracted from the papers themselves. (In the manual extraction, we may have introduced errors, although we did numerous checks to ensure that we did not. Errors may also have been introduced due to misinterpretation of the language in the case of certain older papers.) An auxiliary file contains tables of the points plotted in each figure so that future researchers can easily use these data sets.

Below, we outline our analyses of the data sets. Further details about each data set and our manipulations of the data are given in the Appendix.

For each data set, for each subject, for each stimulus (i.e., a given phase on a given frequency and contrast of the flickering background), we represented the probe thresholds in two different forms:

- The *linear-relative probe threshold* is the probe's threshold luminance (as a function of phase) on the flickering background divided by the probe's threshold luminance on the steady-state background. The steady-state background is typically a nonflickering (0 Hz) background at the same time-averaged luminance as the flickering background.

- The *log-relative probe threshold* is described in either of two equivalent ways. First, it equals the logarithm (to the base 10) of the linear-relative probe threshold. Second, it is equivalently equal to the log of the probe's threshold luminance on the flickering background minus the log of the probe's threshold luminance on the steady-state background.

For a given data set, we then averaged the linear form (respectively, log form) of the thresholds across all the subjects (which varied from 1 to 5) before plotting or further processing the results.

In this article, we generally show the *log*-relative probe thresholds plotted as a function of phase for different frequencies rather than the *linear*-relative probe thresholds. The plots of linear-relative probe thresholds are very much like the plots of log-relative probe thresholds except that the vertical axis has a monotonic but uneven stretching. For one data set, the first introduced in the Results section, we do plot both the log- and linear-relative probe thresholds (Figure 3, top-left and bottom-left panels).

For two of the data sets ("B" for Boynton et al., 1961, and "m" for Maruyama & Takahashi, 1977), the steady-state thresholds were not available to us. Thus, we could not use the above relative measures.
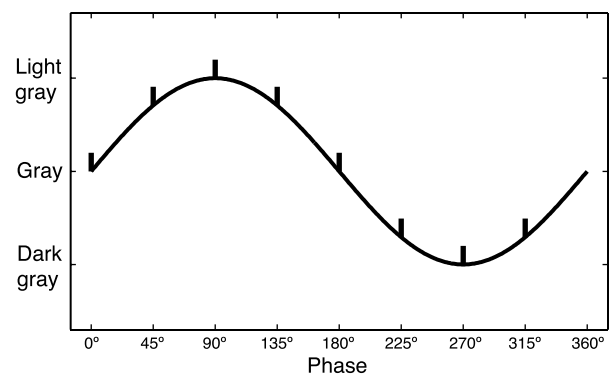

Figure 2. An example of a temporal profile used in the probed-sinewave paradigm. The sinusoidal curve shows the luminance of the flickering background. The eight short vertical lines show eight times (*phases*) at which an incremental probe might be presented (although on a given trial, only one probe would be presented). In the conventions used in this article, a phase of 0 deg indicates a positive zero crossing, and so forth.

	Boynton, Sturr, and Ikeda (1961)	DeMarco, Hughes and Purkiss (2000)	Hood, Graham, von Wiegand and Chase (1997)	Maruyama and Takahashi (1977)	Shady (2000, pp. 39-74)	Shady (2000, pp. 75-91)	Shickman (1970)		Snippe, Poot and van Hateren (2000)	Snippe, Poot and van Hateren (2004)	Wolfson and Graham (2001a)		Wu, Burns, Elsner, Eskew and He (1997)
Authors													
Symbol	B	D	H	(star) m	Y	# †	K	k	S ‡	(gray circle)	W	w	U ‡
Fig. #s here	5,9	4,5	4,5,8,10,11	7,9	4,5,8,10,11	3,10,11	4,5,8,10,11	6,7,8,10,11	4,5,8,10,11	8	4,5,8,10,11	6,7,8,10,11	4,5,8,10,11
Equipment	Optical (Max. view)	Optical (Max. view)	Optical (Max. view)	Optical (Max. view)	Optical (Max. view)	Optical (Max. view)	Optical (Max. view)		Optical (Max. view)	Optical (Max. view)	Computer monitor (free viewing)		Optical (Max. view)
Light source ("Color")	32-cp v-filament automobile spotlight	570 nm Filter and xenon arc lamp	627 nm LEDs "red"	Glow-mod. "white"	660nm LEDs "red"	450nm LEDs "blue"	Bkd.= tungsten Probe = glow-mod.		563nm LEDs "green"	563nm LEDs "green"	CRT "gray"		594 nm He-Ne laser
Mean luminance (L_0)	127 td	741 td	250 td	31.4 td	250 td	0.5 scot td	1280 td	2560 td	7500 td	7500 td	52 cd/m2 (~ 250 td)		2300 td
Waveform of background flicker	Squarewave	Sinusoidal	Sinusoidal	Sinusoidal	Sinusoidal	Sinusoidal	Sinusoidal		Sinusoidal	Sinusoidal	Sinusoidal		Sinusoidal
Contrast of background flicker	100%	63%	57%	28.60%	57%	57%	50%	25%	80%	1-80%	57%	28.50%	100% peak in Gaussian W=5.5 cyc
Cycles of background flicker before probe	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous		Continuous	Continuous	>2.5 s		Several cycles
Probe duration	3 ms	100 ms	10 ms	1 ms	10 ms	10 ms	2 ms		7.5 ms	7.5 ms	13 ms		2 ms
Polarity of probe	Increment	Increment	Increment	Increment	Increment	Increment	Increment		Increment	Increment	Decrement		Increment
Probe diameter	1 deg	Same as background	1 deg (2 deg total)	0.86 deg	1 deg (2 deg total)	1 deg (2 deg total)	2 deg		46 min	46 min	1 deg (1.5 deg total)		1.6 deg
Edge of probe	Sharp	Sharp	Gradual 0.5 deg	Sharp	Gradual 0.5 deg	Gradual 0.5 deg	Sharp		Sharp	Sharp	Gradual 0.25 deg		Sharp
Background diameter	2 deg in dark surround	2 deg in dark surround	18 deg in dark surround	1.72 deg in dark surround	15 deg in dark surround	15 deg in dark surround	22 deg in dark surround		17 deg in dark surround	17 deg in dark surround	7 deg (10 deg total) in L_0 surround (17x13 deg)		9.5 deg in dark surround
Edge of background	Sharp	Sharp	Sharp	Sharp	Sharp	Sharp	Sharp		Sharp	Sharp	Gradual 1.5 deg		Sharp
Psychophysical method	YN staircase	Adjustment	YN staircase	Adjustment	YN staircase	YN staircase	Adjustment		YN	YN	YN staircase		2AFC staircase
No. of phases tested	Many	8	8	Many	8	8	17		4-12	4 (then averaged)	8		9
Phases intermixed?	No	No	No	No	No	No	No		No	No	Yes		Yes
Steady-state thresholds measured?	No	Yes	Yes	No	Yes	Yes	Yes		Yes	Yes	Yes		"Control" between flicker bursts
No. of S's averaged here	1	3	2	2	3	2	2		1-3	1-4	3	5	2
Fig. # in paper showing data plotted here	2	6	4	4 and 9	9	26	5 and 6		1	9 and unpublished data	2		3

† In Figure 3 left-panels, the number sign in the middle of the symbol here is replaced by numbers that refer to the frequency of the flickering background.

‡ In Figure 4 lower-right panel, the letter in the middle of the symbol is replaced by a number to distinguish multiple sets of data from the same authors. The key in that panel shows the frequency associated with each symbol's number.

Table 1. Summary of the data sets plotted in this article. Columns are in alphabetical order based on the first author of each study. The symbols shown here (second row of table) are used in the subsequent figures (third row of table) in this article. *Lowercase letters* inside symbols indicate data collected with low background-flicker contrasts. *Uppercase letters* indicate data collected with high background-flicker contrasts. *Black symbols* with white letters—"D," "k," "K," "S," "w," "W," and "U"—indicate data sets (collected under photopic conditions) in the "majority" group. *White symbols* with black letters—"H" and "Y"—indicate (photopic) data sets with results that differ systematically from the majority. *Star symbols* with letters "B" and "m" are from two early studies for which we do not have steady-state thresholds and which are not clearly part of either group. *Gray symbols* are used in two different ways. The gray diamonds are used for a scotopic data set. The gray circles are used for (photopic) phase-averaged data collected at many background-flicker contrasts.

Summary measures: dc level and peak-to-trough distance (log and linear forms)

Four summary measures were computed for each data set (for each frequency and contrast of background) where it was possible.

1. The *linear form of the dc level* was calculated as follows: The linear-relative probe-threshold values at the different phases were averaged. Values above 1.0 show thresholds that are elevated relative to the steady state. Values less than 1.0 show facilitation.

2. The *log form of the dc level* was calculated as follows: The log-relative probe-threshold values at the different phases were averaged. Values greater than 0 show thresholds that are elevated relative to the steady state, and values less than 0 show facilitation.

3. The *linear form of the peak-to-trough distance* was calculated as follows: The difference was taken between the maximum and minimum values (across phase) of the linear-relative probe thresholds.

4. The *log form of the peak-to-trough distance* was calculated as follows: The difference was taken
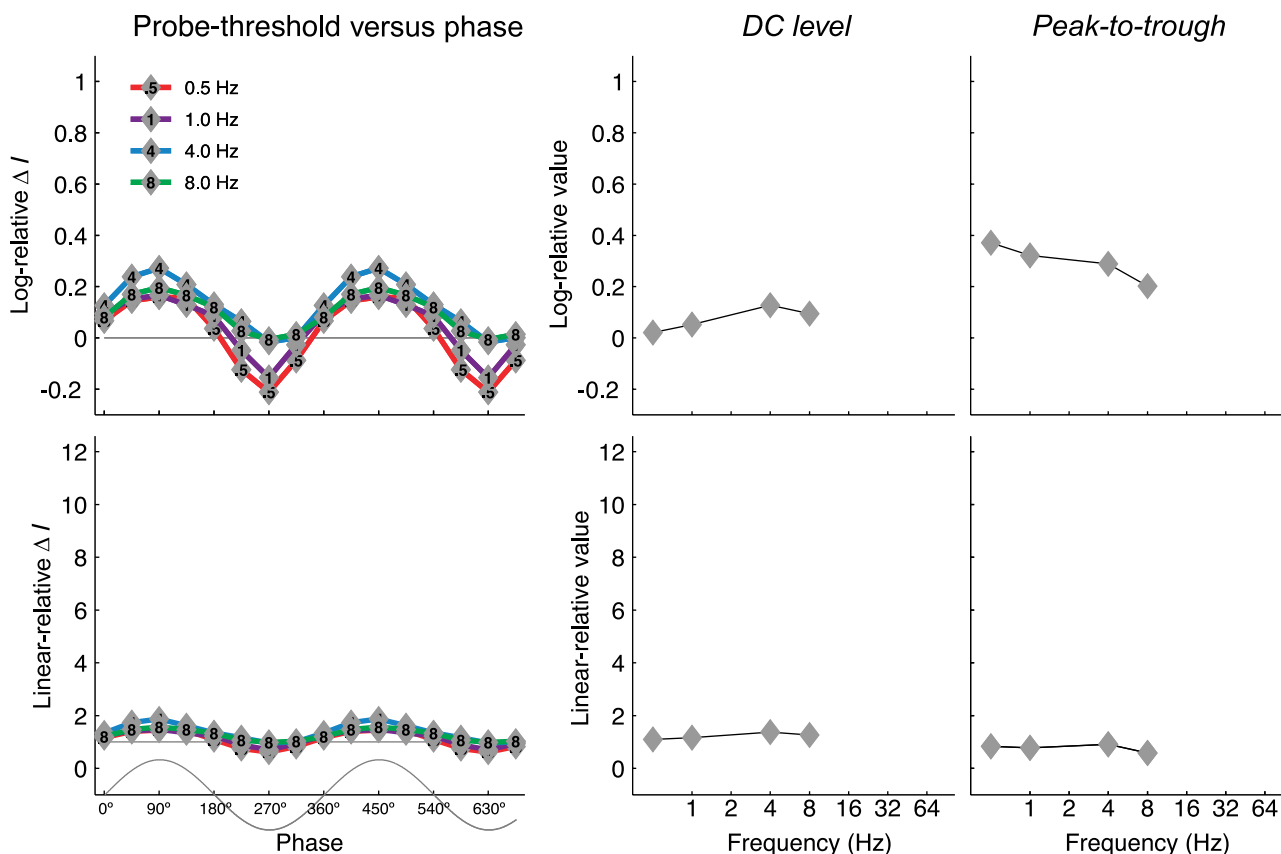
Figure 3. Results from Shady (2000, pp. 75–91) collected under conditions in which the rod system was isolated. The left panels show the log (top-left panel) and linear (bottom-left panel) relative probe thresholds plotted as a function of phase (with respect to the flickering background). The flickering background is shown by the sketch at the bottom of the lower panel. In these panels, the numbers in the middle of the symbols refer to the frequencies of the flickering background, and different colors also distinguish the different frequencies. Each cycle of thresholds has been repeated twice for clarity. The middle and right panels show the four summary measures (different panels) from these results. All four summary measures are plotted as a function of the flickering background's temporal frequency. The middle panels plot *dc level* (relative probe thresholds averaged across phase). The right panels plot *peak-to-trough* distance (difference between the peak and the trough of the relative probe-threshold curves). The top panels show the *log* forms of these summary measures; the bottom panels show the *linear* forms.

between the maximum and minimum values (across phase) of the log-relative probe thresholds.

For the "m" and "B" data sets, where the steady-state threshold values were not known to us, only one of these four summary measures (the log form of the peak-to-trough distance) could be computed.

Results

Rod-system results

We will start our examination of the probed-sinewave data by looking at the one data set collected to isolate the rod system (Shady, 2000, pp. 75–91). These rod-system results appear simpler than the other results. Thus, the rod-system results serve as a convenient introduction.

Figure 3 shows the log-relative probe threshold (top-left panel) and the linear-relative probe threshold (bottom-left panel) plotted as a function of phase (of the probe with respect to the flickering background). A caricature of the luminance of the flickering background is sketched (at an arbitrary height and arbitrary amplitude) along the horizontal axis. The axes here have been drawn so that this figure is comparable with the subsequent figures in this article. The four curves with data points are for the four different frequencies of the flickering background: 0.5, 1, 4, and 8 Hz.

The probe-threshold curves in the left column of Figure 3 are all roughly sinusoidal in shape, and they are in phase

with the sinusoidal flickering background. But note that there is a slight shift upward to higher thresholds as the frequency of the background increases (particularly at the troughs of the curves) and a slight distortion of shape. Shady (2000, pp. 75–91) also collected, though did not show, data at 12 and 16 Hz. At 12 Hz, he reported some threshold fluctuation as a function of phase (though with high variability), but by 16 Hz, the flicker was imperceptible and thresholds were flat at the steady-state level (0 on log relative; 1 on linear relative).

The differences among the curves at different frequencies, minor as they are, show up in the four summary measures shown in the middle and right columns of Figure 3. The layout of the four summary measures in this figure is the same as in subsequent figures. The two rows give the linear and log forms, and the two columns give the dc level and peak-to-trough distance types of the four summary measures defined in the Methods section. Although the curves in three of these four panels are quite flat, the log relative peak-to-trough (top-right panel) curve is definitely decreasing. As will be shown below, this same pattern shows up more dramatically in the results from the cone system.

Cone-system results

All the data sets in Table 1, other than the rod-system results already plotted in Figure 3, were collected under photopic conditions biased toward showing the behavior of the cone system. The probe-threshold versus phase curves from these cone-system data sets are given in three different figures to make it easier to see the similarities and differences among these data sets. Figures 4 and 6 show log-relative probe thresholds as a function of phase for higher (≥50%) and lower (<50%) contrasts of the flickering background, respectively. Different panels in Figures 4 and 6 show different frequency ranges as labeled at the top of each panel. For two studies ("B" and "m"), we could not determine the steady-state thresholds, and therefore, their probe-threshold versus phase results are shown in a separate figure (Figure 9) plotted as log probe threshold rather than relative log probe threshold.

In addition to these three probe-threshold versus phase figures, Figures 5 and 7 show the four summary measures. Figure 5 shows the four summary measures from the high-contrast background results shown in Figure 4 and also the one calculable summary measure from the "B" results in Figure 9. The format of this figure is like that of the middle and right columns of Figure 3. Analogously, Figure 7 shows the summary measures for the low-contrast background data sets of Figure 6 and the "m" data set from Figure 9. Below, we first describe the high-contrast results, then the lower-contrast results, and finally the two studies for which we do not know the steady-state thresholds are described.

High-contrast flickering backgrounds (Figures 4 and 5)

Figure 4 shows the results for the higher contrast (≥50%) flickering backgrounds. At low frequencies of the flickering background (top panels of Figure 4), although there are certainly differences among the curves, the curves in all the studies are quite similar in terms of height and shape. They all have clearly defined narrow troughs at 270 deg, with a broad range of phases producing near-peak values. (They seem to differ slightly in "bumpiness," which we will discuss later.)

In the two panels representing middle frequencies (middle panels of Figure 4), two of the curves are different from the rest, those plotted with white symbols ("H" for Hood et al., 1997, and "Y" for Shady, 2000, pp. 39–74). All the data sets other than "H" and "Y" show much the same shape as at the lower frequencies; that is, they show a narrow trough at 270 deg and a broad range of phases yielding near-peak values. The "H" and "Y" data sets, on the other hand, never show these features in this middle-frequency range. Instead they show the following:

1. In the middle-left panel, the "Y" curve is much higher than the rest, and rather than having a clearly defined trough at 270 deg, there is a range of phases from 135 to 270 deg, all producing values near the trough. Also, rather than having a broad peak, this curve has a narrow peak occurring at a phase of 0 deg (equivalently 360 deg). The height and peak-to-trough distance of the "H" curve is much lower than that of "Y," but the general shape of the curve is similar with a broad trough at much the same phases as those of the "Y" results and a peak at 0 and 45 deg.

2. In the middle-right panel, both the "H" and "Y" curves are higher and flatter than the rest of the curves. The 8-Hz "Y" curve has a broad peak and a broad trough (sharing some similarities with the 4-Hz "H" curve in the middle-left panel). The 8-Hz "H" curve peak and trough are phase-shifted relative to all the other curves. (This change in the phase of the peak and trough was one of the features noted by Hood et al., 1997, and explicitly modeled by Sherman & Spitzer, 2000, in their attempt to model these data.)

The above description emphasizes the differences between the "H" and "Y" results and the "majority results" in this middle-frequency range, but it is important to note the similarities among curves as well: *The linear probe-threshold versus phase curves for all the studies rise vertically in these middle frequencies in such a way that even at the minima of the curves, the thresholds are distinctly elevated above the steady-state threshold.* For example, in the 8- to 13-Hz range (middle-right panel), the lowest threshold is still 0.3 log units above the steady-state threshold! This attribute of the data is important and we use it to help discriminate among models in the Discussion section.

Now, let us look at the higher frequency range, shown in the bottom panels of Figure 4.

1. On the basis of the results in this range, one might not want to say that "Y" and "H" are definitely different (or

definitely not different) from the majority results. For one thing, there is only one curve from each of these studies in this whole frequency range. For another thing, while that 16-Hz "H" curve looks distinctly different from all the others in the bottom-left panel, the 16-Hz "Y" curve does not. (On the other hand, that 16-Hz "Y" curve [bottom-left panel] does look quite a bit like the 8-Hz "H" curve [middle-right panel], including showing the phase shift mentioned above.)

2. There is an overall trend with temporal frequency in this high-frequency range, which is the opposite of that in the low- to middle-frequency range. In this high-frequency range, the probe-threshold versus phase curves fall as frequency increases. To see this, compare the curves in the 16- to 25-Hz range (bottom-left panel) with the curves in the 30- to 40-Hz range (the top three curves of the bottom-right panel) and with the curves in the 50- to 100-Hz range (the bottom five curves of the bottom-right panel). As frequency increases, the curves drop toward the steady-state level (0 on these plots).

Another way of looking at the high-contrast results—and of comparing the "H" and "Y" results with the majority results—is to look at the four summary measures plotted in Figure 5 (from the results shown in Figure 4). Overall, there is an *interaction among the summary measures:* For all the data–sets, three of the four summary measures (top left, bottom left, and bottom right) look very similar and bandpass with values rising to a peak in some middle-frequency range and decreasing on either side. The log peak-to-trough summary measure (top right) is different, it being primarily lowpass.

In the three panels of Figure 5 showing bandpass functions, there is a difference between the "H" and "Y" results and the other data sets: The "H" and "Y" functions peak around 8 Hz, whereas the other data sets' functions peak at substantially higher temporal frequencies (consistent with a peak of about 24 Hz). Thus, the "H" and "Y" results look like the other results shifted to the left (and up a bit). In the fourth panel showing lowpass functions, the difference is not quite as clear. The "H" and "Y" curves are in the same range as the "W" and "K" curves ("W" for Wolfson & Graham, 2001a, and "K" for Shickman, 1970), but all these curves are shifted left (and down a bit) from the "B," "S," and "U" curves ("B" for Boynton et al., 1961, "S" for Snippe et al., 2000, and "U" for Wu et al., 1997). If we just look at the decreasing portions of the curves, then the "H" and "Y" curves again look like the other results shifted to the left. These shifts could mean that effective frequency in the "H" and "Y" experiments was (for some reason) higher than in the other experiments. To put it another way, much of the difference between the "H" and "Y" results and the majority results can be summarized by saying that the effects of the flickering background generally occur at lower temporal frequencies for the "H" and "Y" experiments than those for the majority experiments.

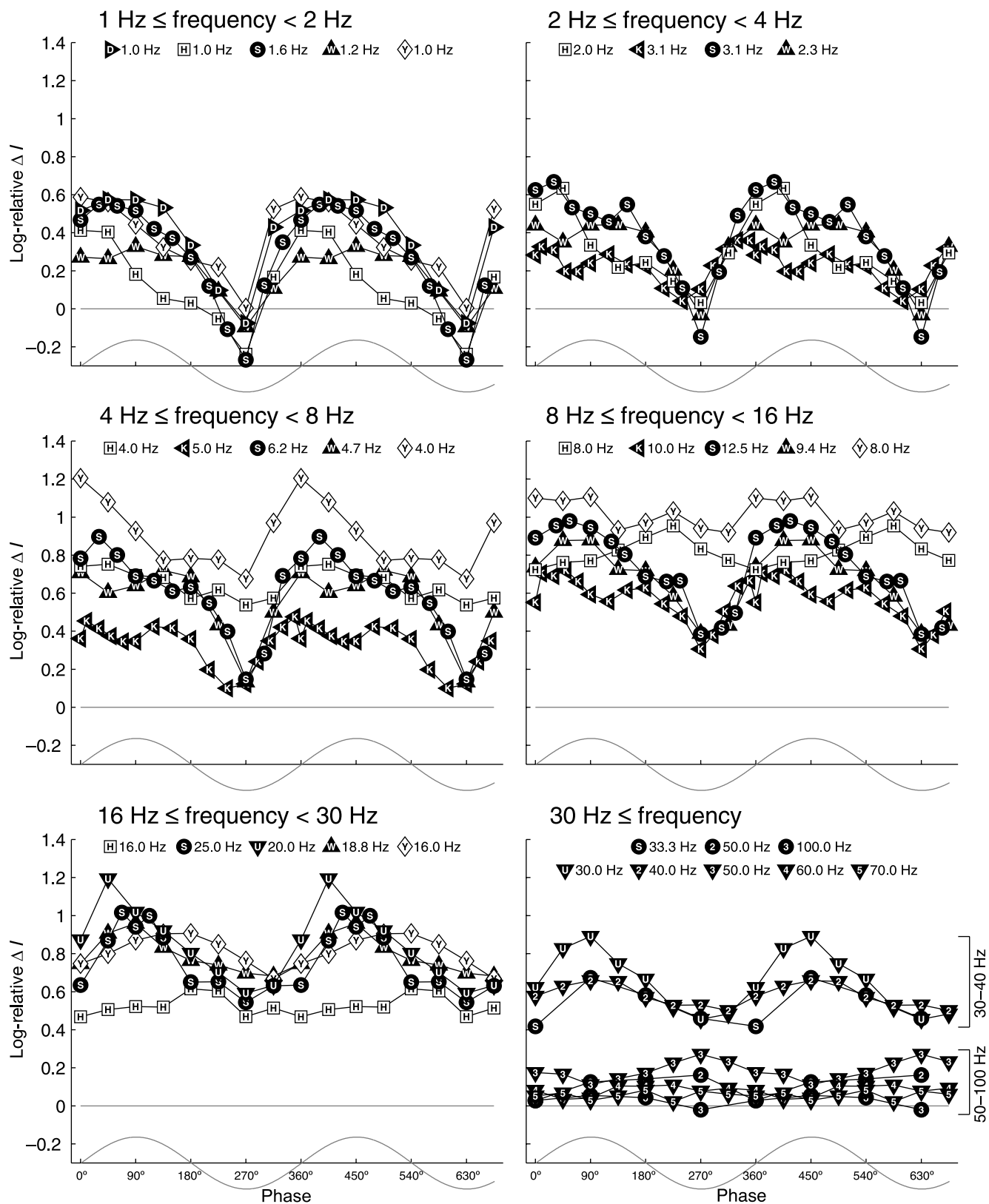### Low-contrast flickering backgrounds (Figures 6 and 7) and the effect of contrast (Figure 8)

Only two data sets from Table 1 used flickering-background contrasts less than 50% (and reported steady-state thresholds), and Figure 6 shows these: "k" and "w" ("k" for Shickman, 1970, and "w" for Wolfson & Graham, 2001a). Note that the lowercase letters indicate low-contrast backgrounds; uppercase letters indicate high-contrast backgrounds. These two data sets are rather similar to one another and quite similar to the majority results at higher contrasts (Figure 4). In particular, at low and medium frequencies (top and middle rows, Figure 6), these two data sets show a distinct, rather narrow, trough near 270 deg while at the one high frequency studied at lower contrast (bottom-left panel, Figure 6), the trough has broadened. The "k" curves (and, to some extent, the "w" curves) show a good deal of bumpiness.

The dc level and peak-to-trough summary measures for these low-contrast background results are shown in Figure 7. They tend to be lower than the results for high-contrast backgrounds (Figure 5). However, the interaction among the four summary measures mentioned above is also seen with these low-contrast background results.

A more complete picture of the relationship between the flickering background's contrast and the dc level is shown in Figure 8, which plots log relative dc level against the contrast of the flickering background (where different groups of frequencies are in different panels). The symbols with letters in them are the same as those used in all the other figures. The gray circles are some data from Snippe et al. (These data were collected in connection with Snippe et al., 2004. These data are not shown in the other figures here because they only reported dc level, not threshold as a function of phase, and because we already have the other complete data set plotted as "S" for Snippe et al., 2000.) Other researchers going back to Shickman (1970)

Figure 4. Results from data sets where the contrast of the flickering background was high (at least 50%). The log-relative probe threshold (vertical axis) is plotted as a function of the probe's phase with respect to the flickering background (horizontal axis). Frequencies are grouped as indicated at the top of each panel. Each cycle of thresholds has been repeated twice for clarity. The flickering background is shown by the sketch at the bottom of each panel. The symbols are those used in Table 1 for the different data sets with some elaboration for the bottom-right panel. The bottom-right panel contains many curves, but these come from only two studies, the Snippe et al. (2000) study shown with circle symbols and the Wu et al. (1997) study shown with downward-pointing triangles. For each study in this panel, one curve is marked by symbols containing the usual letter ("S" or "U"). That is the curve for the lowest frequency shown in this panel for that study. The symbols on the curves for higher frequencies are numbered sequentially from 2 onward. The legend in the bottom-right panel gives the actual values of the frequencies connected with each number.

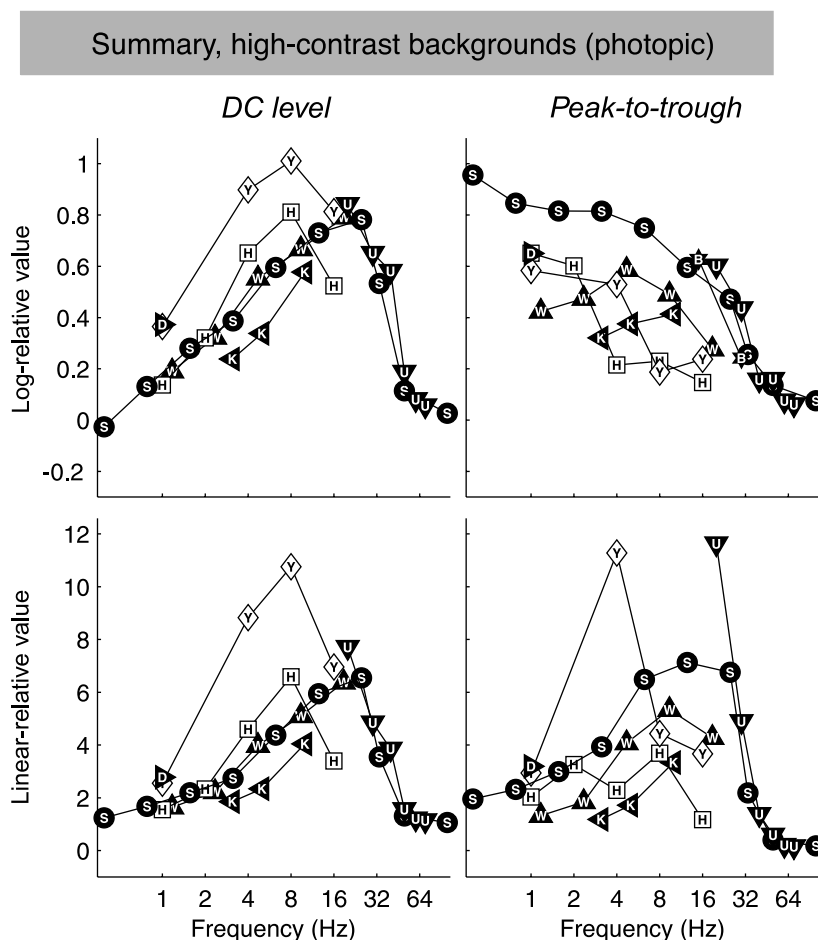## Summary, high-contrast backgrounds (photopic)

Figure 5. Plots of the four summary measures (different panels) for the high-contrast flickering-background results shown in Figure 4 (and the "B" data set from Figure 9). All four summary measures are plotted as a function of the flickering background's temporal frequency. Left panels plot *dc level* (relative probe thresholds averaged across phase). Right panels plot *peak-to-trough* distance (difference between the peak and the trough of the relative probe-threshold curves). The top panels show the log forms of these summary measures; the bottom panels show the linear forms. The symbols used for the data points are those identifying the columns in Table 1.

have looked at the contrast level of the background flicker, but this is the most complete set of data we have seen. The dc level increases as the contrast of the flickering background increases, and this increase is more dramatic at higher frequencies of the flickering background (although, presumably, the dc level would drop if the frequency continued to increase, as seen in the top-left panel of Figure 5 on the right end of the *x*-axis). As in the earlier figures, the results in Figure 8 that do not fit the general trends are the "H" and "Y" results.

Studies without steady-state thresholds (Figure 9)

Finally, the remaining two data sets from Table 1 are shown in Figure 9. For these two data sets ("B" and "m"), we do not have steady-state threshold levels, so the results cannot be plotted relative to steady state, but instead the thresholds themselves are shown. The one summary measure that can be plotted from each is the log peak-to-trough measure, and it is plotted in Figure 5 for the "B"

results (collected with a high-contrast background) and in Figure 7 for the "m" results (collected with a low-contrast background).

The "B" results on a high-contrast 15-Hz background (top-right panel, Figure 9) look very similar to those of the several studies in Figure 4 using high-contrast stimuli in the 16- to 25-Hz range (bottom-left panel), for example, compare it with the "U" curve at 20 Hz. Also, the "B" results on a high-contrast 30-Hz background (Figure 9, bottom-right panel) fit in well with the other studies' curves collected in this range (Figure 4, bottom-right panel, upper three curves). The one summary measure for the "B" results (Figure 5, top-right panel) fits reasonably well with the majority data. However, there are only two frequencies in the "B" study, there is no reported steady state, and there is only one subject, thus, not too much should be made of these data.

The "m" results on low-contrast backgrounds at 2 and 10 Hz (Figure 9, left panels) are not easy to categorize either with "H" and "Y" or with the majority results. This categorization
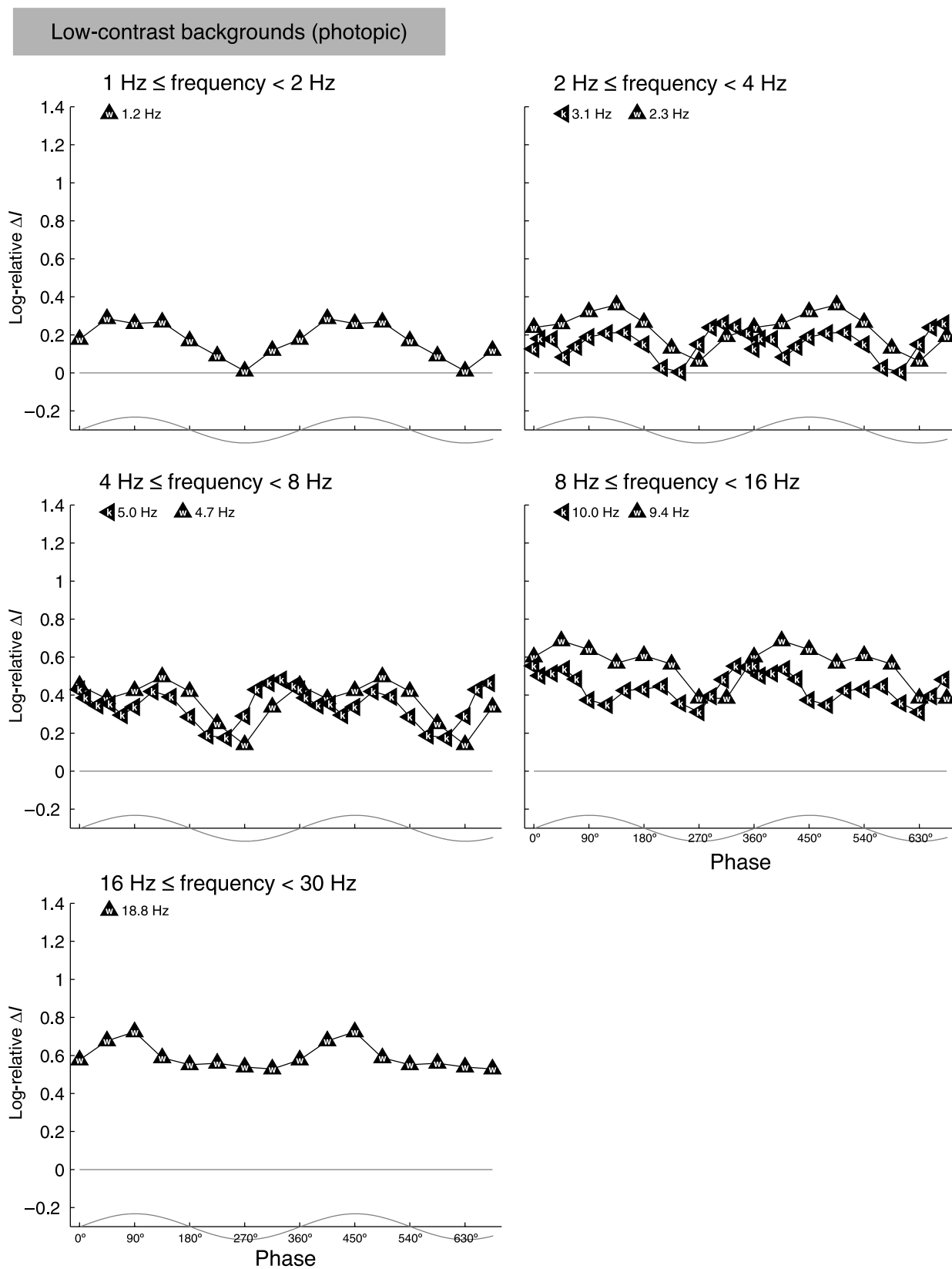
Figure 6. Results from data sets where the contrast of the flickering background was less than 50% plotted as log-relative probe threshold (vertical axis) versus phase (horizontal axis). Conventions are the same as those in Figure 4. The sketch at the bottom of each panel shows the flickering background to have less amplitude than that in Figure 4 to indicate the lower contrast. Lowercase symbol letters are also used to indicate lower contrast.
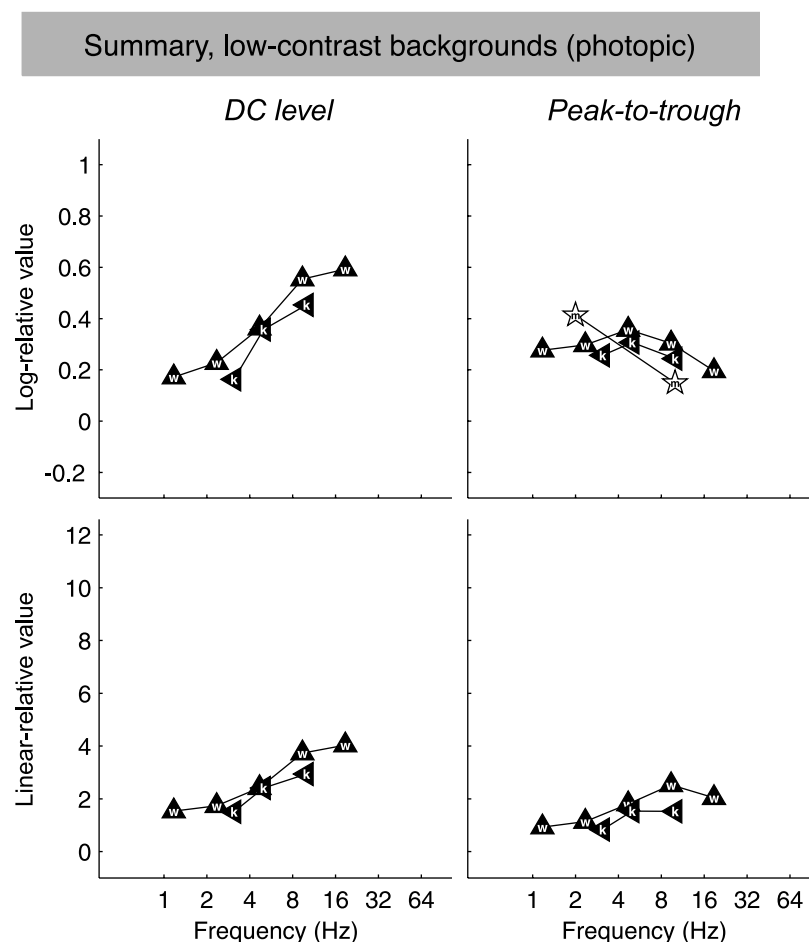
Figure 7. Plots of the four summary measures for the low-contrast flickering background results shown in Figure 6 (and the "m" data from Figure 9). Conventions are the same as those in Figure 5.

is hard because there is so little "m" data, because the data were collected at low contrasts where "H" and "Y" did not collect data, and because the steady-state level is missing. There is a hint in the one summary measure for the "m" results (Figure 7, top-right panel) of a slight left shift on the frequency axis relative to the two majority studies in that panel. Notice that the "m" data were collected with a very low mean luminance. We will return to this later.

Differences in bumpiness

As mentioned above, there is bumpiness in some of the probe-threshold versus phase curves. At least some of that bumpiness may come from the intrusion of a frequency-doubled process, that is, a process that would produce two peaks and two troughs in one cycle. This bumpiness, seen most strongly in the "K," "k," and "m" curves (Figures 4, 6, and 9), does not have any obvious correlate in the summary measure plots (Figures 5 and 7).

Note that all of the low-contrast probe-threshold versus phase curves show bumpiness at some frequencies. We are not sure if there is much to be made of this because (1) the low-contrast "k" curves show a great deal of

bumpiness but so do the high-contrast "K" curves; (2) the low-contrast "w" curves show some bumpiness, but the high-contrast "W" curves show about the same amount; and (3) the low-contrast "m" curves show bumpiness but they did not collect high-contrast results.

Discussion

This Discussion is divided into three sections. First, we consider the explanatory value of the experimental variables outlined in Table 1. Then, we discuss what the experimental results suggest about visual processes. Lastly, we discuss the models that have been compared with probed-sinewave results.

Explanatory value of the experimental variables in Table 1

Are there particular differences in the experimental setups that can account for the differences we see in the
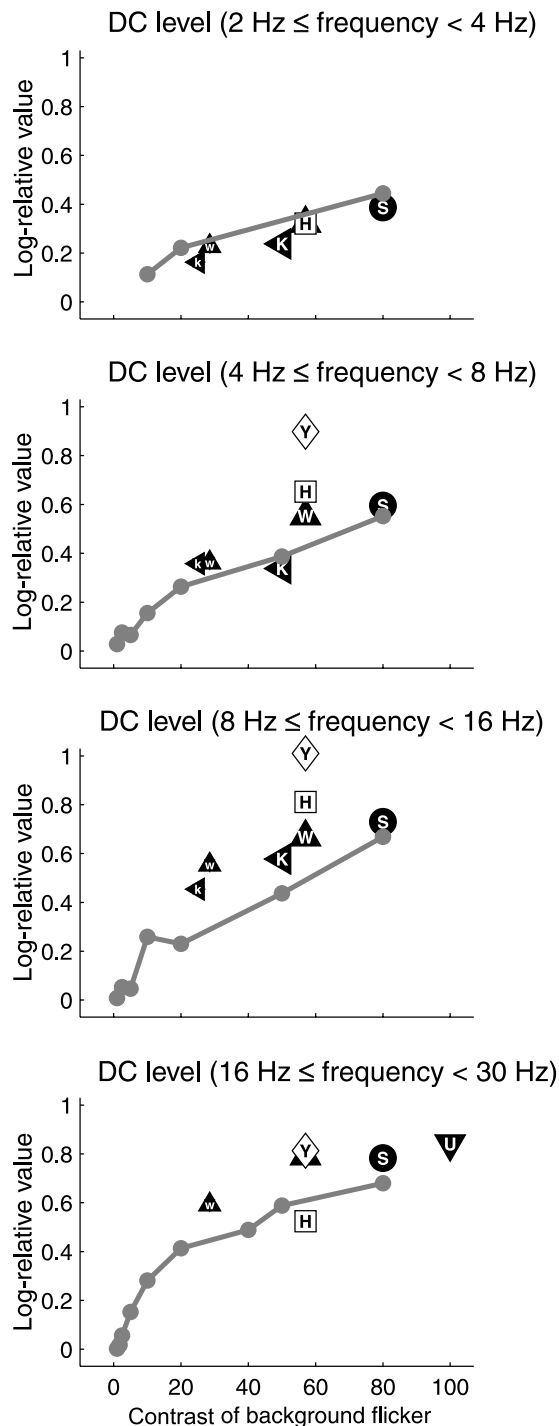
Figure 8. Plots of the log form of the dc level summary measure versus the contrast of the background flicker, grouped by frequency of the background. Symbols with letters show the same data as the top-left panels in Figures 5 and 7. Gray-circle points are at frequencies of 3.125 (top panel), 6.25, 12.5, and 25 Hz (bottom panel). The 25-Hz data are from Snippe et al. (2004); the results for the other three frequencies are previously unpublished data provided by Snippe. See Appendix for details.

probe-threshold versus phase curves across data sets? We will consider each row in Table 1, but, first, the short answers. As reported in the Results section, we find two differences:

1. The "H" and "Y" results diverge from the rest of the high-contrast results, particularly at middle frequencies of the flickering background. There are a number of differences between the "H" and "Y" experimental setups and the rest of the experimental setups, but we will dismiss all but one below. Namely, the *Light source* ("*Color*") used in the "H" and "Y" experiments was long wavelength. None of the other experiments isolated long wavelengths. "H" and "Y" used long-wavelength light to isolate the long-wave cone system. We will discuss this more in the next section.

2. There is a greater bumpiness in the probe-threshold versus phase curves for some data sets (in particular, "K," "k," and "m"). There is no experimental parameter in Table 1 that clearly differentiates these experiments from the others.

Remember that, although there are these differences among the results, *there is also a great deal of similarity*. This is remarkable given the extreme differences in experimental conditions listed in Table 1 and analyzed next.

a. *Equipment:* The general type of stimulus display used—whether an optical system or a CRT—did not matter. To see that, compare the "W" results in Figure 4 with the other results in that figure and the "w" results in Figure 6 with the other results in that figure.

b. *Light source* ("*Color*"): The spectral content of the light varied across studies. At low and high frequencies of the flickering background, this does not seem to matter (because all the data sets are generally in agreement). However—as mentioned above and discussed more below—at mid frequencies, the results of "H" and "Y" do diverge from the rest and these are the only results collected with long-wavelength light.

c. *Mean luminance* (L_o): The mean luminance across the data sets varied dramatically, but this made little difference when probe-threshold results were plotted relative to steady state as shown in Figures 4 and 6.

Wu et al. (1997) conducted a within-study comparison of mean luminance. They compared data collected at mean luminances between 580 and 9,100 td at 30 and 50 Hz. Their results show that the shapes of the probe-threshold curves change little across these mean luminances. They did find that the *absolute* dc level increased with increasing mean luminance but so did the steady-state threshold. Thus, their results show—as do the results reported here—that *relative* probe threshold is quite constant across mean luminance.
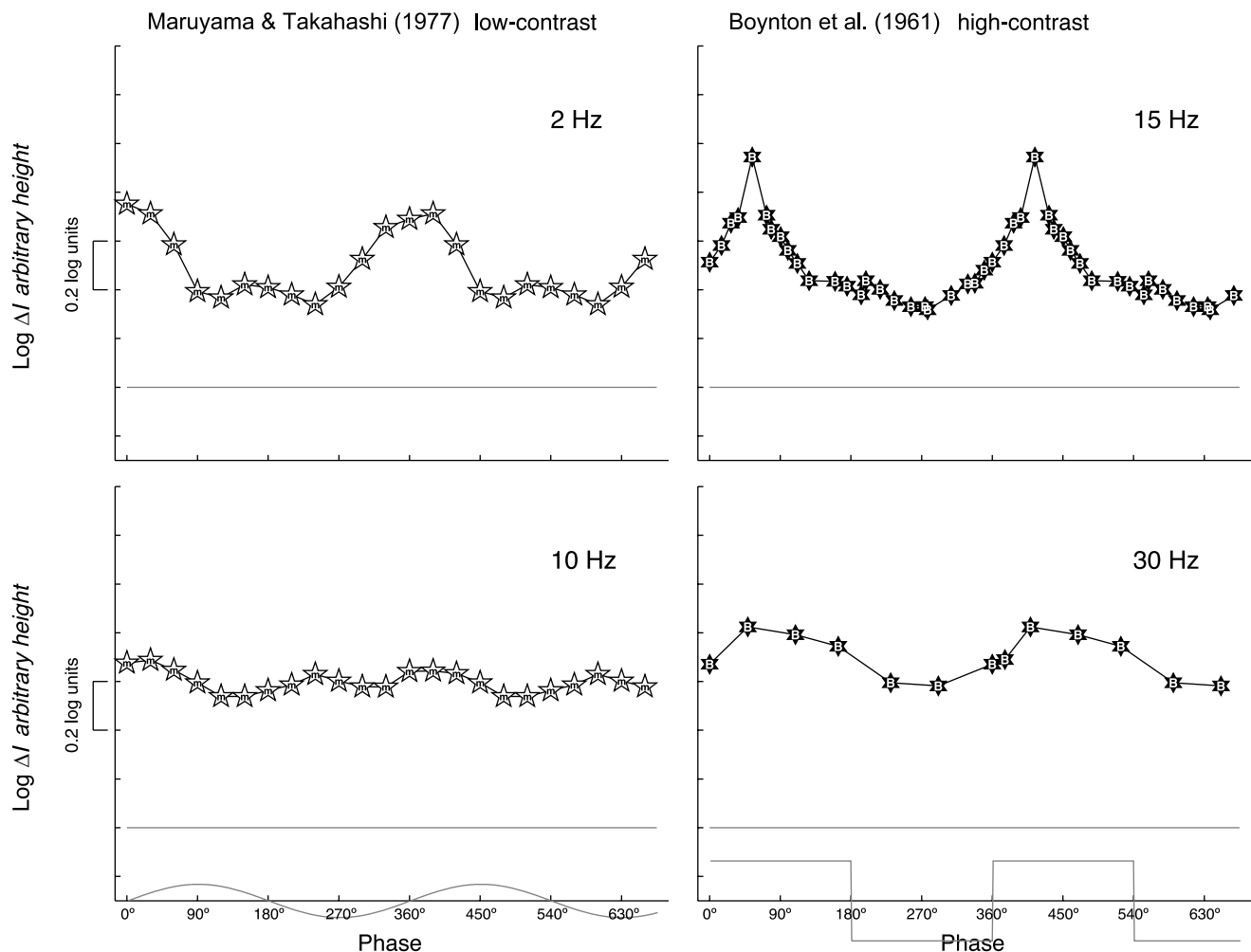
Figure 9. Log probe-threshold curves as a function of phase and frequency for the two studies where we do not know the steady-state thresholds. Hence, the vertical placement of the curves on the *y*-axis is arbitrary because it cannot be determined from the reported data. Conventions are the same as those in Figures 4 and 6.

It is *not* the case that mean luminance differentiates the "H" and "Y" experiments from the majority results because "H" and "Y" used the same mean luminance (250 td) as the "W" experiment. We raise this point because Snippe et al. (2000) noted that, at middle frequencies, their results ("S") diverge from the "H" results, and they suggested three possible reasons for the divergence: (1) mean luminance level, (2) soft-edge versus sharp-edge probes, and (3) light source color. We have just dismissed the first reason, and we will dismiss the second reason below. We agree with the third reason, and we will discuss this further below.

It is the case that the "m" data—which we feel unable to group with either the "H" and "Y" data or with the majority data because such data reflect aspects of both—were collected at a lower mean luminance

than any other data. The "m" data were collected at 31.4 td. Maybe very low mean luminances matter? We will also discuss this further below.

d. *Waveform of background flicker:* With the exception of the "B" results, all of the results were collected using sinusoidal flicker. The intent of this article is to compare results collected with sinusoidal flicker, and we only included the "B" results for historical reasons. Note that Maruyama and Takahashi (1977) collected probe thresholds on 2- and 10-Hz squarewave-flickering backgrounds. We do not plot those results in this article, and they are too low in terms of both luminance and frequency to reasonably compare to the "B" results.

e. *Contrast of background flicker:* The contrast of the background flicker does not alter the general shape of

the probe-threshold versus phase curves much (compare Figure 6 to Figure 4) but does decrease the dc level and peak-to-trough distance (compare Figure 7 to Figure 5). The relationship of dc level to background contrast is shown more finely in Figure 8.

There is a tendency for the probe-threshold versus phase curves with greater bumpiness to come from the low-contrast experiments, but, as discussed above, we do not think this parameter explains the bumpiness.

f. *Cycles of background flicker before probe:* Most of the experiments used continuous flicker. Those experiments that did not use continuous flicker ("W," "w," and "U") do not have different results than the other comparable experiments. This is not surprising because Snippe et al. (2004) and Wolfson and Graham (2000) found that, within 40 ms of the start of the background's flickering, threshold reaches the level obtained when the background has been flickering continuously.

g. *Probe duration:* The probe duration varied from 2 to 13 ms (for all but one of the data sets) and did not differentiate the results. The one data set with a probe duration outside this range is that of DeMarco et al. (2000) who used a 100-ms probe and a 1-Hz flickering background. The results with this 100-ms probe are very similar to the other comparable results (compare the "D" curve in the top-left panel of Figure 4 with the other curves).

A more straightforward comparison of probe duration can be found in DeMarco et al. (2000). In that article, they compared probe durations of 12, 25, 50, and 100 ms at 1 Hz. They found minimal change in curve shape although they did find some *absolute* threshold elevation for the shortest duration probe. They only collected steady-state thresholds for the 100-ms probe condition; thus, we cannot comment on *relative* thresholds. The fact that they only collected steady-state thresholds for the 100-ms probe is precisely why those data are plotted as "D" in this article.

h. *Polarity of probe:* Increment probes and decrement probes do have slightly different results as analyzed in both Wolfson and Graham (2001a) and DeMarco et al. (2000). However, the differences are slight, relative to the differences we are considering in this article, as can be seen, for example, by comparing the "W" results in Figure 4 with the other results in that figure.

i. The next four rows in Table 1 are about spatial characteristics of the stimuli. We were interested in the spatial characteristics because Snippe et al. (2000) had listed soft-edge versus sharp-edge probes as a possible reason for the differences between "S" and "H" results. In addition, we were interested in the spatial-contrast characteristics of the stimuli because,

in related experimental paradigms, they can make a huge difference (e.g., Spehar & Zaidi, 1997).

Although the spatial characteristics differ dramatically across data sets, they do not do so in a way that can account for the differences among the results. The *Probe diameter* varies from about 0.5 to 2 deg. The *Edge of probe* is sometimes sharp and sometimes gradual, and while the "H" and "Y" results were both collected using a gradual edge, so were the "W" results, indicating that this difference in spatial characteristic does not produce the differences in the results. The *Background diameter* varies dramatically—from 2 to 22 deg—as does the relationship of the probe diameter to the background diameter (from coincident to a very small probe in a very big background). Generally, the background is in a dark surround and the *edge of background* is sharp, but a mean luminance surround and a smooth edge, as used to collect the "W" data, do not change the pattern of results.

j. *Psychophysical methods:* The psychophysical methods include variants of adjustment, detection, and forced-choice procedures, but this does not seem to matter much. While it is true that the probe-threshold versus phase curves with the greatest bumpiness were collected using adjustment methods, the "D" results were also collected using adjustment and do not show bumpiness. Thus, we do not think this parameter explains the bumpiness.

k. The *No. of phases tested* differs greatly, but this does not seem to affect the results. It is true that "K," "k," and "m" tested at more phases than most, and that their probe-threshold versus phase curves show more bumpiness. However, "S" also tested some frequencies at more phases but does not show a great deal of bumpiness. Hence, we do not think that testing at more phases would reveal significantly more bumpiness in the other curves.

l. Several miscellaneous aspects of the experiments are listed in the last few rows of Table 1, but none of these made any difference in the pattern of results. Sometimes, the *Phases are intermixed* and sometimes they are not. The *Steady-state thresholds were measured* for all the results in Figures 4 and 6. The results without steady-state thresholds are plotted separately in Figure 9. Finally, the *No. of S's averaged here* varies but is generally at least 2. Results from single subjects were only included when we considered the results unique enough to warrant inclusion.

Implications for visual processing
Why long wavelengths (and mean luminance) might matter

As we said before, much of the difference between the "H" and "Y" results and the majority results can be
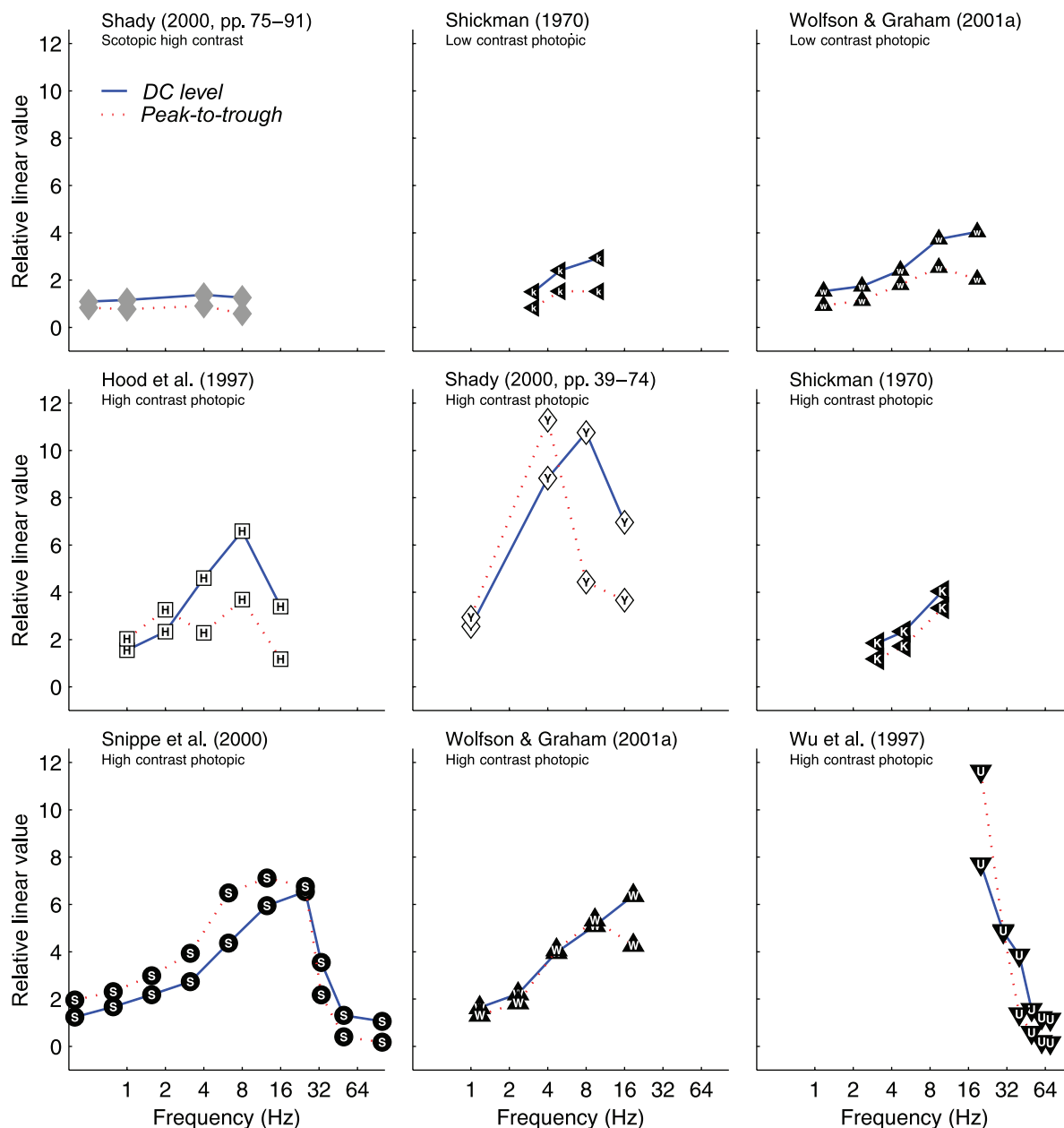
Figure 10. The linear relative peak-to-trough (dotted red lines) and dc level (solid blue lines) summary measures plotted versus temporal frequency of the flickering background. Different data sets are shown in different panels. All the data sets from Table 1 are shown (except for the "D" data set that has only one frequency, and the "B" and "m" data sets where we do not know the steady-state threshold level and thus cannot compute the dc level measure). Notice that, as frequency gets higher, the dc level (solid blue lines) generally gets higher relative to the peak-to-trough distance (dotted red lines).

summarized by saying that the effects of the flickering background generally occur at lower temporal frequencies for the "H" and "Y" experiments than for the majority experiments. The best evidence for this is shown in Figure 5: For all four summary measures, the curves for "H" and "Y" are different than the others, particularly in the middle

frequency range. However, if one were allowed to shift the "H" and "Y" curves to the right, they would look more like the other curves.

A possible reason for this difference can be found in the one experimental parameter in Table 1 that distinguishes "H" and "Y" from the majority: the spectral content of the
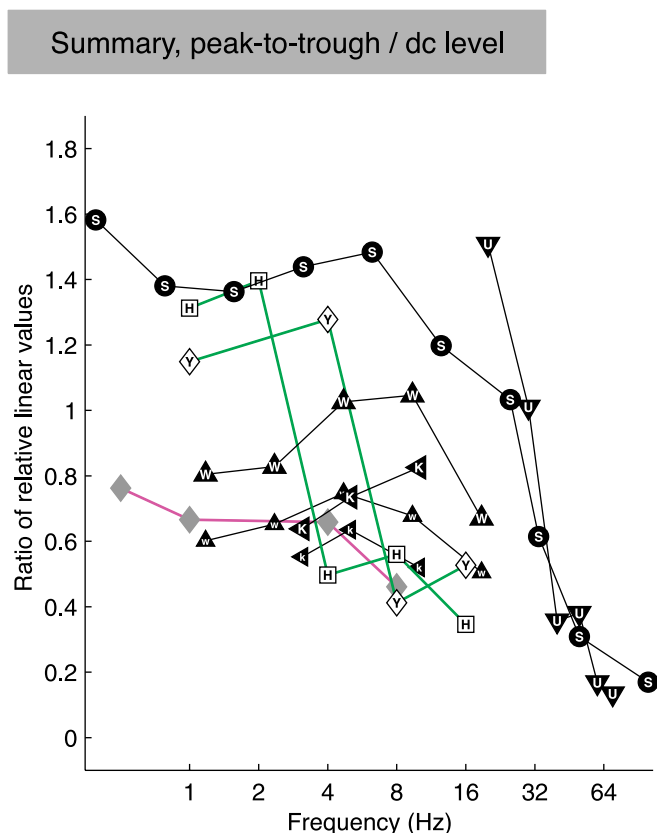
Figure 11. The vertical axis gives the ratio of the linear relative peak-to-trough summary measure divided by the linear-relative dc level summary measure (plotted in Figure 10). Each curve represents a different data set (with the usual symbols). Scrutiny reveals the following: (1) all the data sets except "K" show a high-frequency decline, (2) the curves for "H" and "Y" (green lines) have this same shape but are translated left on the frequency axis, and (3) the curve for the rod-system results (pink line) also shows this same trend.

stimulus. "H" and "Y" used long-wavelength light to isolate the long-wave cone system. We speculate that using long-wavelength light may emphasize the contributions from parvocellular/chromatic pathways, whereas the shorter wavelength or broadband light used by the other experimenters may emphasize the contributions from magnocellular/achromatic pathways. In general, many previous psychophysical and physiological studies suggest that the parvocellular/chromatic pathways are responsive within a range of relatively low temporal frequencies, whereas the magnocellular/achromatic pathways are responsive out to the highest visible temporal frequencies. Thus, it may not be too surprising that the flickering background's effects in the "H" and "Y" experiments (using long-wavelength light) occur in a lower temporal-frequency range than those for the majority experiments (even the "W" ones using the same mean luminance but broadband light). Of course, without further empirical work, one could not be certain of this explanation. It is,

however, certainly plausible. Indeed, Snippe et al. (2000) briefly discussed the possible roles of pathways in their experiments, leaning toward magnocellular mediation for their results (plotted here as "S").

The summary measures for low-contrast backgrounds are shown in Figure 7. In the top-right panel of Figure 7, there is a *hint* that the one summary curve available for "m"— the log peak-to-trough distance—is shifted to lower temporal frequencies than the other curves in that panel. This feature of the "m" data is somewhat like that of the "H" and "Y" data (in Figure 5). Remember that the "m" study used a much lower mean luminance (31.4 td) than that used by the majority studies. In general, psychophysical and physiological results show a move of contrast sensitivity toward lower temporal frequencies as mean luminance decreases (holding spectral content constant). Thus, if there is a leftward shift in the "m" results relative to the majority results, one would suspect that lower mean luminance (while holding spectral content constant) was its cause. Again, without further empirical work, one could not be certain.

Relationship of dc level to peak-to-trough distance

The effects of the flickering background frequency on the peak-to-trough distance and the dc level can be compared directly as shown in Figure 10 for all the data sets in Table 1 (except for "D," "B," and "m"). Each panel shows the linear form of these two summary measures plotted against background frequency for a particular data set. Notice that, as frequency increases, the linear dc level summary measure (blue solid line) rises relative to the linear peak-to-trough summary measure (red dotted line). This is true for all the data sets except for "K."

Yet another way to picture the relationship between the dc level and peak-to-trough summary measures is to plot the ratio of the two as was done in Figure 11. The nine curves in Figure 11 correspond to the nine different panels in Figure 10. The vertical axis shows the linear peak-to-trough summary measure divided by the linear dc level summary measure. This ratio declines at high frequencies. Note that, plotted in this manner, the descending portion of the "H" and "Y" results (green lines) appear, again, to be shifted left relative to the other results. The rod-system results, plotted with a pink line, resemble the photopic results, although the effect is much smaller.

One way to think about the effect of temporal frequency on the ratio of peak-to-trough to dc level might be as follows: perhaps there are two rather separate "reasons" or "processes" underlying the effects on the dc level and the peak-to-trough distance. Perhaps the process producing elevation in dc level is relatively more sensitive at higher frequencies than is the process producing elevation in peak-to-trough distance. There are many possible processes. For example, Wolfson and Graham (2001b) found that probe thresholds were monocularly mediated at 1.2 Hz, but at 9.4 Hz, some of the dc level elevation appears to be due to

dichoptic processing. This switch in processing is in the appropriate temporal frequency range for the difference noted here.

Model of light adaptation dynamics compared to probed-sinewave results

There are four current models (some with several versions) that have attempted to predict results from probed-sinewave studies of light adaptation. In each of the four subsections below, we will discuss one of the models. The paragraphs starting with *Biggest Success* and *Biggest Failure* are relative to the whole set of models at this point in time. *Other* contains information about other features of the model's ability to account for known results. *Conclusion* sums up our current feeling about the attractiveness of this model.

Model 1: Merged models of Graham and Hood (1992)

The Merged Models of Graham and Hood (1992) combined pieces of light adaptation models from both the "periodic tradition" and "aperiodic tradition." Models from one tradition (periodic or aperiodic) were found to predict results from their own tradition but *not* from the other tradition. The Merged Models were created to predict results from *both* traditions.

Biggest Success: The Merged Models successfully predict (Graham & Hood, 1992; von Wiegand, Hood, & Graham, 1995) older psychophysical results from both traditions.

Biggest Failure: For the probed-sinewave paradigm, these models fail (Hood et al., 1997). In particular, these models do *not* show dc level overtaking peak-to-trough distance as temporal frequency increases. Indeed, the predicted trough threshold never rises above the steady-state threshold even at 16 Hz. Wu et al. (1997) also reject a more extended version of some of the MUSNOL models considered by Hood et al. (1997).

Other: One of the models, Merged 2, does a number of things rather well. In particular, it can predict some of the bumpiness through a tendency toward frequency doubling (Hood et al., 1997). It can predict some phase shifts (Hood et al., 1997). It predicts some of the measured effect of varying the duration of the flickering background before the probe (Wolfson & Graham, 2000). It can also predict some slight differences between increment and decrement probes (Wolfson & Graham, 2001a).

Conclusion: The "biggest failure" is *critical* as it concerns a commonly found feature of the data.

Model 2: Wilson (1997) modified by Hood and Graham (1998)

The Wilson (1997) model was modified by Hood and Graham (1998) to predict probed-sinewave results. The model is based on retinal physiology, making it rather complex.

Biggest Success: This model does predict that, as frequency increases, the dc level should overtake the peak-to-trough distance and, in line with this, the trough threshold does rise above the steady-state threshold (Hood & Graham, 1998; Wolfson & Graham, 2000, 2001a).

Biggest Failure: There is substantial difficulty in reconciling the components of the model with known physiology. It was the known physiology that inspired the model, and without such a fit between physiology and model components, the complexity of the model makes it less useful for understanding the phenomena and has no particular justification. Specifically, there is a problem in the fit between the component in the model that does the most work in the "biggest success"—namely, the push–pull circuit necessary to predict the dc shift correctly—and the fact that the push–pull circuit occurs in primate cortex but not earlier than cortex as far as anyone knows (Hood, 1998; see also Hood & Graham, 1998; Snippe et al., 2000; Wolfson & Graham, 2000, 2001a for further discussion). This is also a problem because dichoptic versus monocular experiments at 1.2 and 9.4 Hz show that dichoptic presentation of the flickering background and the probe eliminates much of the elevation in dc level (Wolfson & Graham, 2001b). Snippe et al. (2000, p. 459) mention another problem with the retinal physiology assumed by Wilson's model and calculate a new retinal prediction from that model that could be checked.

Other: The model can predict some bumpiness in the probe-threshold versus phase curves in large part due to the presence of "on" and "off" mechanisms (Hood & Graham, 1998). It can predict some phase shifts (see Hood & Graham, 1998, who also show the components of this for low and high frequencies). It captures some aspects of the effects of varying the duration of the flickering background before the probe (Wolfson & Graham, 2000). It does a reasonable job of predicting the differences between increment and decrement probes (Wolfson & Graham, 2001a). Moreover, it correctly predicts the aperiodic traditions' probe-flash results and the periodic tradition's temporal-frequency-dependent effect of mean luminance (Wilson, 1997).

This is the only one among the four models discussed here to have distinct magnocellular and parvocellular pathways.

Conclusion: This model is still in contention as far as a computational model goes. The physiological justification for the push–pull component is currently weak. Without such justification, the complexity of the components makes this model less attractive to us in terms of the utility of the model to increase our understanding.

Model 3: Dahari and Spitzer (1996) and Sherman and Spitzer (2000)

The Dahari and Spitzer (1996) and Sherman and Spitzer (2000) model is based strongly on retinal ganglion cell

physiology. It was aimed at the "H" results partly because those results showed such a strong phase effect (a phase shift) for peaks and troughs.

Biggest Success: This model is unique among the models in being able to correctly predict the huge phase shift of the trough and peak with frequency shown in the "H"(and "Y") results. However, because this shift is *not* seen in the majority results, this can be viewed as a limited success.

Biggest Failure: As with the Merged Models of Graham and Hood (1992), this model does *not* show dc level overtaking peak-to-trough distance as frequency increases. Consistent with this, the trough threshold never rises above the steady-state threshold even at 16 Hz.

Other: There seems to be no bumpiness in the predictions. The model's predictions for the duration of flicker before the probe and the differences between increment and decrement probes have not been computed. As it contains what are supposed to be retinal components, it does correctly predict that the effect should be primarily retinal at low frequencies but will miss the cortical component at middle frequencies.

Conclusion: The "biggest failure" (as with the Merged Models) is critical.

Model 4: Snippe et al. (2000, 2004)

The Snippe et al. (2000, 2004) model is based on three abstract components not tied to physiology: a divisive light adaptation process, a subtractive light adaptation process, and a contrast-gain control process. The contrast-gain control accounts for the change in dc level with frequency seen in the results. (The dynamics of the contrast-gain control are examined in the 2004 paper.) This model was compared to different probed-sinewave results with emphasis on the "S" results.

Biggest Success: Like the Wilson model and unlike the other two, this model correctly predicts that the dc level overtakes the peak-to-trough distance at high frequencies and, in line with this, the threshold at the trough rises well above steady-state level.

Biggest Failure: The published predictions do not show much bumpiness. However, there are some plateaus in the published predictions. Further, with changes in parameters, the model *can* predict more bumpiness and even frequency doubling.

Other: It does predict some shift of the phase of peak and trough, although not as large as seen in the "H" (and "Y") results. The effect of the duration of flicker before the probe is explicitly measured and built into the contrast-gain component in the 2004 study; thus, it is now correctly predicted by the model. The small increment versus decrement difference may well be predictable using this model (see discussion in Wolfson & Graham, 2001a, pp. 1130–1131).

This model's abstract components (two kinds of light adaptation—subtractive and divisive—plus an explicit contrast-gain control) allow a deeper understanding more easily of what is important in the model and what is not. Of course, any relationship to the physiology is then lost and needs to be separately considered. This lack of attempted correspondence between physiology and components means that this model is neutral with regard to predicting the dichoptic experiment.

Conclusion: This is the most attractive model to us at present because it predicts most of the effects very well, at least for the majority results, and it might well be modifiable to predict the "H" and "Y" results.

Summary

Overall, the data across laboratories are remarkably similar given the substantial differences in stimulus and experimental parameters. The shapes of the probe-threshold versus phase curves are quite similar for most data sets (within a range of frequencies, within a range of contrasts, see Figures 4, 6, and 9). As the contrast of the flickering background increases, the dc level increases (see Figure 8). There is an interaction among the four summary measures: as the high-contrast background's frequency increases, the linear dc level, the log dc level, and the linear peak-to-trough difference are bandpass, whereas the log peak-to-trough difference is generally lowpass (see Figure 5). Although not as dramatic, this pattern of results is present for the low-contrast backgrounds shown in Figure 7 (and is also suggested by the results from the rod system shown in Figure 3). In addition, as shown in Figures 10 and 11, whatever produces the shift in dc level becomes increasingly more important relative to whatever produces the shift in peak to trough as frequency increases.

In terms of the differences among data sets, we find that the "H" and "Y" results, particularly at middle frequencies, are somewhat different than the majority results. The "H" and "Y" results appear shifted to lower frequencies than the majority results. Our best guess about the experimental condition that causes the differences is the long-wavelength light used by "H" and "Y" to isolate the long-wave cone system. This could have increased the contributions from the parvocellular/chromatic pathways, and this could interact with the frequency of the flickering background.

Finally, we briefly discussed the four models (and their variants) that have been challenged by probed-sinewave results. The one that seems most attractive to us currently is the Snippe et al. (2000, 2004) model.

Appendix: Details of the data sets

Each of the data sets used in this article is outlined in Table 1. Below are further notes about each data set. In particular,

we have tried to state all of the assumptions and choices we made for each data set to put them into a common format for the figures in this article. The general process went like this:

1. We obtained the raw threshold values either from the authors directly or by taking the values from the published figures.
2. We transformed the phases, if needed, to our convention shown in Figure 2.
3. For each data set, for each subject, we changed absolute thresholds ("ΔI") to relative thresholds ("relative ΔI"). Relative means relative to the steady state that is defined below for each data set but typically was threshold measured on a steady (0 Hz) background at the same mean luminance as the flickering background. (This was only possible, of course, when the steady state was reported.)
4. For each data set, for each subject, we computed both the relative linear thresholds ("linear relative ΔI") and the relative log thresholds ("log relative ΔI").
5. Within a data set, we then averaged across subjects.
6. Finally, we plotted the probe-threshold versus phase curves (Figure 3, left column, and Figures 4, 6, and 9) or calculated the summary measures (Figure 3, right columns, and Figures 5, 7, 8, 10, and 11).

One last general note. We use only one "canonical" set of data per set of authors in the probe-threshold versus phase plots. If the authors have published other relevant data, such data will be noted below. The most glaring example of authors with a great deal of data, of which only one set is used in this article, is ourselves. Although there are some differences across the data of Wolfson and Graham (2000, 2001a, 2001b), these differences are small relative to the differences being examined in this article. We felt that inclusion of all the data would overwhelm the figures without adding enough additional insight.

Boynton et al. (1961, their Figure 2)

The Boynton et al. (1961) data were taken from the published figures.

They measured increment probe thresholds on 15- and 30-Hz flickering backgrounds for one subject. The flickering background's waveform in this case was squarewave, not sinewave. These were the only data we include that were collected with squarewave flicker. We include these data because they are historically significant.

To process the data, we converted their phase values to correspond to ours, calling the increment step in the squarewave 0 deg and the decrement step 180 deg. No absolute threshold levels were given, although their plots indicated the distance corresponding to 1 log unit; hence, we chose an arbitrary 0 level and measured relative to that. At 30 Hz, threshold was measured over three cycles of the background, but we only used the data from the first cycle here (although all three cycles look quite similar).

(Not included here are Boynton et al., 1961, data collected on a very dim 30-Hz flickering background. The background was so dim that it was not perceived as flickering, but the shape of the probe-threshold versus phase curve in that condition was quite similar to the 30-Hz data described above and plotted in Figure 9.)

DeMarco et al. (2000, their Figure 6)

The DeMarco et al. (2000) data were obtained directly from the authors.

They measured increment and decrement probe thresholds (at 1 Hz). We only use their increment data here. Because they only reported steady-state data for their 100 ms-duration probe, we used those data (rather than the 50-, 25-, or 12-ms data they collected). Their phase values were shifted 180 deg relative to ours; thus, we shifted the values appropriately. They measured probe threshold at each of eight phases (0, 45, …, 315 deg). We used the steady-state level from the zero crossing, although they measured it for each phase. Probe thresholds were plotted as "Probe Threshold (ΔTds)," which we transformed to log and linear relative ΔI for each of the three subjects.

(Not included here are the DeMarco et al., 2000, decrement probe data collected at 1 Hz, nor are the increment and decrement data collected with other probe durations. Also not included here are their increment and decrement data collected with a Gaussian waveform background.)

Hood et al. (1997, their Figure 4)

The Hood et al. (1997) data were obtained directly from the authors.

They measured increment probe threshold at 1, 2, 4, 8, and 16 Hz. Their eight phase values were the same as ours. Steady state was measured on a steady background at the same mean luminance as the flickering background. Probe thresholds were given as "probe threshold (td)," which we transformed to log and linear relative ΔI for each of the two subjects.

(Not included here are Hood et al., 1997, data collected at 6, 10, and 12 Hz for a single subject. Also not included here are their data collected at 0, 1, and 4 Hz with a lower mean luminance and higher flicker contrast. Finally, not included here are their data collected at 1 Hz for a green probe on a green background.)

Maruyama and Takahashi (1997, their Figures 4 and 9)

The Maruyama and Takahashi (1977) data were taken from the published figures.

They measured increment probe threshold at 2 and 10 Hz. We converted their phase values to correspond to ours. (This was slightly complicated because the number of phases was not constant across or within frequency, nor was it the same

across or within subject. Further, the 10-Hz data were so physically small on the paper that precise phases were hard to determine. Thus, we binned phases into 30 deg bins.) No absolute threshold levels were given, although their plots indicated the distance corresponding to 0.3 log units; thus, we chose an arbitrary 0 level and measured relative to that. Multiple plots of data were given for each subject at each frequency of the flickering background. We averaged the data within subject, yielding one set of log ΔI values for each of the two subjects at each of the two frequencies.

(Not included here are Maruyama and Takahashi, 1977, data collected with squarewave and sawtooth flickering backgrounds. We only use their sinewave background data.)

Shady (2000, his Figures 9 and 26)

The photopic data (Shady, 2000, pp. 39–74) were obtained directly from the author. The scotopic data (Shady, 2000, pp. 75–91) were taken from the published thesis figures.

Under photopic conditions, he measured increment probe thresholds at 1, 4, 8, and 16 Hz. Under scotopic conditions, he measured increment probe thresholds at 0.5, 2, 4, and 8 Hz. His phase values were the same as ours. Steady state was measured on a steady background at the same mean luminance as the flickering background. Probe thresholds were reported as linear "relative ΔI," which we transformed to log relative ΔI for each of the subjects.

Some of these data were presented in Shady (1999).

(The scotopic data we plot in this article was collected with a mean luminance of 0.5 scot td and flickering background contrast of 57%. Not included here are additional scotopic data collected under the following conditions: (1) 0.5 scot td and 20% contrast, (2) 0.5 scot td and 100% contrast, (3) 0.003 scot td and 57% contrast, (4) 0.126 scot td and 57% contrast, and (5) 2.000 scot td and 57% contrast.)

Shickman (1970, his Figures 5 and 6)

The Shickman (1970) data were taken from the published figures.

He measured increment probe thresholds at 3.1, 5.0, and 10 Hz at background contrast levels of 25% and 50%. Multiple probe thresholds were measured each 20 deg (using the same phase convention that we used); we averaged the multiple measurements. Steady-state levels were given for steady backgrounds at the minimum and maximum luminances; we averaged these values to generate a single steady-state value. Values were reported as "test energy (td-sec)," which we converted to log and linear relative ΔI for each of the two subjects.

(Not included here are the following Shickman, 1970, probe-threshold versus phase data. Data for one subject at 3.1, 4.0, 5.0, 6.2, 7.7, and 10 Hz with a mean luminance of

640 td and a background contrast of 99%. Data for another subject at 3.1, 5.0, 6.2, and 10 Hz with a mean luminance of 64 td and a background contrast of 99%.)

Snippe et al. (2000, their Figure 1)

The Snippe et al. (2000) data were obtained directly from the authors.

They measured increment probe thresholds at 10 different frequencies (0.39, 0.78, 1.56, 3.13, 6.25, 12.5, 25, 33, 50, and 100 Hz) of the flickering background. In Figure 4, we plotted the data collected with flickering backgrounds greater than 1 Hz. In Figure 5, we plotted all the frequencies. Their phase values were the same as ours. They measured probe threshold at four phases (0, 90, 180, and 270 deg) at all the frequencies of the flickering background and additional phases for one subject at some of these frequencies. Steady state was measured on a steady background at the same mean luminance as the flickering background. Probe thresholds were given as "test detection threshold (Td)" in their article. We transformed the data to log and linear relative ΔI for each of the (up to three) subjects.

Snippe et al. (2004, their Figure 9 and unpublished data)

The Snippe et al. (2004) data were obtained directly from the authors.

They measured increment probe thresholds at four frequencies of the flickering background (3.125, 6.25, 12.5, and 25 Hz). Note that only the 25-Hz data are plotted in Snippe et al. (2004); the authors kindly provided us with additional data. They measured probe threshold at four phases (0, 90, 180, and 270 deg) and averaged the data to get dc level. Steady state was measured on a steady background at the same mean luminance as the flickering background. The contrast of the flickering background varied from 0% (steady state) to 80%; the number of these contrast levels differed across frequency. We transformed the dc level data to make them relative for each of the (up to four) subjects.

(Not included here are Snippe et al., 2004, data—averaged across phase—measured at various times before the background began to flicker, after the background stopped flickering, and before/after a change in the contrast of the background flicker. See also additional phase-averaged data in Snippe and van Hateren, 2003.)

Wolfson and Graham (2001a, their Figure 2)

The Wolfson and Graham (2001a) data were already in our files.

We collected increment and decrement probe thresholds at five frequencies (1.2, 2.3, 4.7, 9.4, and 18.8 Hz) at eight phases of the flickering background (those in Figure 2) for two contrast levels (28.5% and 57%). In this article, we

used the decrement data (rather than the increment data) because these data were more complete. Steady state was measured on a steady background at the same mean luminance as the flickering background. The data were reported as "relative log ΔI" for each of three subjects with the high-contrast background and five subjects with the low-contrast background.

(As stated above, we did not include our Wolfson and Graham, 2001a, increment data here. Nor have we included data about the perceived probe-polarity from that same paper. Nor have we included the Wolfson and Graham, 2000, data collected at 1.2 and 9.4 Hz using variable amounts of flicker before and after the probe to investigate timing. Nor have we included the Wolfson and Graham, 2001b, data collected at 1.2 and 9.4 Hz under multiple viewing conditions: monoptic, dichoptic, and binocular.)

Wu et al. (1997, their Figure 3)

The Wu et al. (1997) data were taken from the published figures.

They measured increment probe threshold for six frequencies of the flickering background (20, 30, 40, 50, 60, and 70 Hz) at each of the nine phases (spaced at 45 deg) using the same phase conventions that we used. The stimulus temporal profile was Gaussian windowed such that, when the flickering started, it was at a low background contrast that then increased and then decreased. Probe threshold was measured during the cycle at the center of the Gaussian window. They did not measure thresholds on a steady-state uniform background but instead measured "control" thresholds on a uniform background in between bursts of flicker. These control thresholds varied with background flicker frequency, being greatest at the lowest frequency they measured (20 Hz) and quite low at the highest frequencies (60 and 70 Hz). Here, we took the minimum control thresholds as an estimate of the true steady-state threshold. Thresholds were reported as "Test threshold (td)," which we converted to log and linear relative ΔI for each of the two subjects.

(Not included here are Wu et al., 1997, data collected at 30 and 50 Hz at other mean luminances—580, 1,150, and 9,100 td—and at other background contrasts—25% and 50%.)

Acknowledgments

Commercial relationships: none.
Corresponding author: S. Sabina Wolfson.
Email: sabina@psych.columbia.edu, nvg1@columbia.edu.
Address: Columbia University, Department of Psychology, New York, New York, 10027, USA.

References

Boynton, R. M., Sturr, J. F., & Ikeda, M. (1961). Study of flicker by increment threshold technique. *Journal of the Optical Society of America, 51,* 196–201.

Dahari, R., & Spitzer, H. (1996). Spatiotemporal adaptation model for retinal ganglion cells. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 13,* 419–435. [PubMed]

DeMarco, P. J., Jr., Hughes, A., & Purkiss, T. J. (2000). Increment and decrement detection on temporally modulated fields. *Vision Research, 40,* 1907–1919. [PubMed]

Graham, N., & Hood, D. C. (1992). Modeling the dynamics of light adaptation: The merging of two traditions. *Vision Research, 32,* 1373–1393. [PubMed]

Graham, N., Wolfson, S. S., & Chowdhury, J. (2001). A comparison of light adaptation results from 40 years of the probed-sinewave paradigm [Abstract]. *Investigative Ophthalmology & Visual Science, 42,* abstract #840, p. S157.

Hood, D. C. (1998). Lower-level visual processing and models of light adaptation. *Annual Review of Psychology, 49,* 503–535. [PubMed]

Hood, D. C., & Graham, N. (1998). Threshold fluctuations on temporally modulated backgrounds: A possible physiological explanation based upon a recent computational model. *Visual Neuroscience, 15,* 957–967. [PubMed]

Hood, D. C., Graham, N., von Wiegand, T. E., & Chase, V. M. (1997). Probed-sinewave paradigm: A test of models of light-adaptation dynamics. *Vision Research, 37,* 1177–1191. [PubMed]

Maruyama, K., & Takahashi, M. (1977). Wave form of flickering stimulus and visual masking function. *Tohoku Psychologica Folia, 36,* 120–133.

Shady, S. (1999). The dynamics of light-adaptation of the rod and cone systems [Abstract]. *Investigative Ophthalmology & Visual Science, 40,* abstract #246, p. S46.

Shady, S. (2000). *Comparing the dynamic mechanisms of light adaptation of the rod and cone systems: Empirical investigation and theoretical analysis.*

Doctoral dissertation, Columbia University, New York, NY.

Sherman, E., & Spitzer, H. (2000). Confrontation of retinal adaptation model with key features of psychophysical gain behavior dynamics. In S.-W. Lee, H. H. Bülthoff, & T. Poggio (Eds.), *BMCV 2000, LNCS 1811* (pp. 88–97). Berlin–Heidelberg: Springer-Verlag.

Shickman, G. M. (1970). Visual masking by low-frequency sinusoidally modulated light. *Journal of the Optical Society of America, 60,* 107–117. [PubMed]

Snippe, H. P., Poot, L., & van Hateren, J. H. (2000). A temporal model for early vision that explains detection thresholds for light pulses on flickering backgrounds. *Visual Neuroscience, 17,* 449–462. [PubMed]

Snippe, H. P., Poot, L., & van Hateren, J. H. (2004). Asymmetric dynamics of adaptation after onset and offset of flicker. *Journal of Vision, 4*(1), 1–12, http://journalofvision.org/4/1/1/, doi:10.1167/4.1.1. [PubMed] [Article]

Snippe, H. P., & van Hateren, J. H. (2003). Recovery from contrast adaptation matches ideal-observer predictions. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1321–1330. [PubMed]

Spehar, B., & Zaidi, Q. (1997). Surround effects on the shape of the temporal contrast-sensitivity function. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2517–2525. [PubMed]

von Wiegand, T. E., Hood, D. C., & Graham, N. (1995). Testing a computational model of light-adaptation dynamics. *Vision Research, 35,* 3037–3051. [PubMed]

Wilson, H. R. (1997). A neural model of foveal light adaptation and afterimage formation. *Visual Neuroscience, 14,* 403–423. [PubMed]

Wolfson, S. S., & Graham, N. (2000). Exploring the dynamics of light adaptation: The effects of varying the flickering background's duration in the probed-sinewave paradigm. *Vision Research, 40,* 2277–2289. [PubMed]

Wolfson, S. S., & Graham, N. (2001a). Comparing increment and decrement probes in the probed-sinewave paradigm. *Vision Research, 41,* 1119–1131. [PubMed]

Wolfson, S. S., & Graham, N. (2001b). Processing in the probed-sinewave paradigm is likely retinal. *Visual Neuroscience, 18,* 1003–1010. [PubMed]

Wu, S., Burns, S. A., Elsner, A. E., Eskew, R. T., Jr., & He, J. (1997). Rapid sensitivity changes on flickering backgrounds: Tests of models of light adaptation. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2367–2378. [PubMed]