responsetoreviews

Comments to the Author

General Comments

Wade and Baker have written a very approachable and well-motivated review of the use of VEP measurements of visual contrast. The authors rightly focus on the importance of measuring contrast responses under a wide range of conditions, highlighting the ways in which human evoked responses can be made in informative in diverse contexts. The review should thus be of broad interest.

There are points where the review is overly broad and others that could use more detail. These are described below.

A central part of the review is the role of computational models and normalization models, in particular. The use of the normalization framework for describing the VEP contrast response function (CRF) has been fruitful in terms of mechanistic interpretation and linking to other measurements, such as single-unit physiology and fMRI. This modeling approach assumes that the CRF reflects the activity of a more or less unitary population, and that the population response resembles that of the underlying individual neurons that make up the population.

There are two edge-cases in the literature that suggest a more nuanced approach will be needed. It would be useful to be clearer on the assumptions of the modeling approach and what the significance is for interpretation of cases where the assumptions don't hold.

Two-limbed contrast response functions. Campbell and Maffei (J. Physiol,1970) in the earliest measurements of the VEP contrast response function reported that it sometimes had two "limbs" – one at low contrast and a second one with a steeper slope at higher contrasts. Two-limbed contrast response functions have also been found by Nakayama and Mackeben (Vision Res,1982) in macaque with electrodes on the dura, by Apkarian and Tyler (Vision Res, 1985), Norcia et al., (Vision Res,1989) and Souza et al., (IOVS, 2007), to name a few. As noted, the sigmoidal model assumes that the population response is unitary, but the two-limbed functions suggest that it is a mixture of populations. The two limbs may not be apparent when the function is sampled with only a few points over a wide range of contrast.

Harmonics with different CRFs. Bobak et al., (IOVS, 1986), found that the first harmonic (1f) and second harmonic (2f) of the pattern onset/offset SSVEP had different thresholds but similar slopes once above threshold. Norcia et al., (TVST, 2024) found different thresholds and slope for 1f and 2f responses driven by contrast modulated dynamic noise. This again suggests the presence of multiple underlying populations. The behavior also raises the question of whether to pool harmonics when estimating the response function. This was alluded to in the simulation, but could be expanded on.

The lack of dynamics in the normalization model, as presented should be discussed, the brief mention of Tsai et al. aside. The authors may wish to cite another framework that explicitly models dynamics (e.g. amplitude and phase) in the context of the contrast response function (Zemon and Gordon, Vision Res, 2006) .

Response

> We thank the reviewer for these positive comments. In general this was intended to be an overview in the sense of a 'primer' for researchers coming to the SSVEP for the first time rather than a detailed historical account of the field - hence the inclusion of code examples and relatively domain-agnostic language. However, the reviewer raises some important general issues and we are happy to address them: Specifically, we include references to the idea that SSVEP can engage more than one type of neuron, that the population responses is not trivially derived from what is known about the properties of individual neurons and that fitting with a sigmoidal response function may mask some complexities in the CRF that are more readily seen with denser sampling. We think the Bobak reference is IOVS 1984 and we have also included the Kaestner TVST 2024 reference (thank you for drawing our attention to both of these).

Specific comments

Adaptation effects. Adapted contrast response functions have been measured by Nelson et al., (EEG Journal, 1984) by comparing CRFs measured by sweeping contrast from low to high vs high to low. They found that responses at low contrast were smaller in the "down" sweeps and attributed it to differences in prior adaptation.

> we could not find this reference. We did find an excellent (and sometimes humorous) review of SSVEPs by Nelson (Nelson, J. I., Seiple, W. H., Kupersmith, M. J., & Carr, R. E. (1984). Lock-in Techniques for the Swept Stimulus Evoked Potential. Journal of Clinical Neurophysiology, 1(4), 409–436. doi:10.1097/00004691-198401040-00002) but could not find a mention of adaptation effects.

Other studies also measured adaptation effects on the CRF Heinrich and Bach,

(IOVS, 2001), Ross and Speed (1991, Proc Roy Soc). And the hard-to-find Bach, Greenlee and Buhler (Clin Vis Sci, 1996) paper that modeled the effect within a contrast gain control framework.

> The Ross and Speed paper measures masking using SSVEP but the other adaptation measures are, as far as we can tell, psychophysical. Heinrich and Bach are concerned with transient changes in SSVEP amplitude within a recording which may have an adaptation origin but which manifest as response increases over the first few seconds and seems a little mysterious (we often observe these as well). Bach, Greenlee and Buhler is, however, pretty unequivocal and we now include it. We like the suggestion that the 'increase' in response is attributable to a shift from a square wave response to a more sinusoidal waveform as the apparent contrast of the stimulus is decreased by adaptation-driven gain control. The response power which was distributed across odd harmonics in the original response can now be concentrated in the 1st harmonic after adaptation.

Gain control development. Morrone and Burr (Nature, 1986) used frequency tagging to measure orientation dependence of masking in infants and adults, as did Candy et al. The difference in the two studies is that in adults Morrone and Burr found response gain for cross-oriented maskers but Candy et al found contrast gain. Skoczenski and Norcia used a high contrast 30 Hz dynamic 2D noise masker and variable contrast reversing gratings. They found masking consistent with a contrast gain effect.

> We have clarified that these papers measure different types of gain control

Page 5 bottom. The reciprocal effect of test and masker was first shown by Regan and Regan (1988; cited) for flicker modulation depth and then for gratings by Brown et al., (IOVS, 1999 for dichoptic masking) and then by candy et al., (J. Neurosci, 2001) along with Busse et al. (Neuron, 2009) as cited.

> Thanks - we have included these references

It's good to have cited Regan and Regan paper that introduced the notion of fingerprints in the spectrum being tied to the nature of the non-linearity. Maybe expand their "fingerprint" approach a little more.

> We have expanded a little but are also conscious of limitations to this approach (see also the discussion re unique inputs and system order below). We think there might be some ways to relax these constraints a little by incorporating information about likely transducer functions at different stages (essentially fitting parameters rather than performing a full system identification) but this still requires many inputs.

Figure 4. Oversaturation occurs because the denominator overtakes the numerator at high contrast as it has two terms in it, not the single term in the

numerator. Would be worth showing total power not saturating. Is this a general effect?

> We think this topic is probably worth a more detailed treatment than we can provide here. Broadly, much of the oversaturation effect can be modeled by an increase in saturation and therefore harmonic distortion - power is transferred from the first to higher harmonics as contrast increases and the output becomes more 'square-wave-like'. It seems likely that total power does not saturate (perceptually, there appears to be a monotonic increase in contrast for example and discrimination thresholds increase up to 100%) but the issue is complicated by the fact that the distribution of neurons sensitive to different contrast levels might not be uniform and that neurons might alter their response properties to signal with optimum sensitivity around a particular range. Our goal here is simply to show people new to the field how the CRF is approximately log-linear with contrast.

Equation 3 is a special case where the neurons in the numerator are so selective that they don't "see" the masker except by being influenced indirectly by the gain pool. Putting the test and mask in the numerator and summing them before exponentiating yields robust IM. Weights can be used to reflect the neurons preference for the test vs the mask. Does equation 3 produce IM?

> In the accompanying code we specifically allow the user to alter the way that normalization is computed and many of the (commented) options produce robust IM. We encourage the user to experiment with different configurations to explore their effects. We have updated the text to make this more explicit.

Page 7 Mckeefry et al., (Vis Neuro, 1996) is a good reference for linking harmonics of on/off to transient sustained responses.

> Included

Page 8 figure 5. The extrapolation to zero amplitude method should be attributed to Cambell and Maffei (J Physiol) 1970). Campbell and Kulikowski (J Physiol, 1972) made a detailed comparison to psychophysics.

> We have now included these references

Figure 5 is an oversimplification. As it doesn't take into account how the additive noise mixes with the internal VEP response. This was modeled in Norcia et al., (Vision Res, 1989). The additive noise adds a bias to the measurement, especially if the slope of the response function is shallow. The authors should try to connect the linear regression approach for threshold measurement to the sigmoidal non-linearity, as they are different models of the CRF. In practice the difference at the low contrast end is probably too small to matter and the saturation region is excluded.

> We have now noted that it is important to consider noise for situations

4

where the SNR is low or the slope is shallow and we expand on the point that you have to stick to the part of the CRF where the curve is more or less linear - as we showed explicitly in the figure. The code also includes an explicit 'noise baseline' that can be used to demonstrate changes in the intercept as a function of SNR.

Page 9. The sweep method was first developed by David Regan (Vision Res, 1973) for use in objective refraction.

We have included this reference.

Page 9 "This approach" seems to refer to the sweep VEP, not the method of measuring at multiple contrast levels over a series of trials which was used in the Atkinson/Braddick papers. The sweep method refers to the approach of incrementing or decrementing the stimulus value within a single trial and then averaging the sweeps to get the response function. Regan argued that this was a better way of estimating response functions because amplitude variations over a session are distributed over the whole response function shifting it up or down instead of distorting its shape (Regan (Vison Res, 1975, color coding).

This is a good point but the (simpler) one we are trying to make here is that extrapolating from high stimulus levels down to zero provides a robust estimate of visual sensitivity in subjects where such measurements might otherwise be hard to make. We have altered the text to clarify this.

Page 9 ...chromatic and chromatic contrast as well as stereoscopic depth perception develops earlier than had been supposed previously based on behavioural readouts (Dobson et al., 1978; Norcia & Tyler, 1985b) with both chromatic and achromatic contrast detection reaching near-adult levels by around six months. Needs references for chromatic sensitivity and disparity sensitivity.

The DeVries Khoe and Spekreijse comparison between VEP and psychophysics was done above age 4 years and for acuity not for contrast sensitivity. The large discrepancies between VEP and FPL are in infancy for contrast sensitivity, as well as for acuity. In the realm of contrast sensitivity, the differences are very large, up to a factor of 50. Compare Norcia et al., 1990 Vision Res measurement to any behavioral data. Suggest making this comparison more quantitative.

We have made this more concrete. We have taken out the reference to steroscopic depth (because it is not directly related to contrast). It is hard for us to directly compare the values for contrast sensitivity from FPL vs VEP - the FPL values in particular seem to span quite a range. The figure in Brown 1990 suggests a x2 difference in detection threshold (if we are reading the plot correctly) but we are happy to accept higher values if the reviewer can point us to the relevant FPL literature.

The authors may wish to cite Norcia et al., (NeuroReport, 2000) for an early measurement of suppression in amblyopia using measurements of the contrast

response function in each eye in the presence of a high contrast dichoptic masker in the other. This study showed that amblyopic observers retain suppressive binocular interactions despite severely reduced stereopsis. The dominant eye exerted a greater suppressive effect on the non-dominant eye, suggesting that it maintains a relatively greater number, or more effective, inhibitory contacts with cortical cells. Also may wish to cite recent work by Hou and colleagues using a similar approach, including application of normalization models.

> We think these were both in already but we have explicitly grouped them in the section on amblyopia to make their relevance clearer.

Schade (JOSA, 1956) was the first to measure the contrast sensitivity function of the human visual system as noted by Campbell and Robson (J. Physiol, 1968). Campbell and Green (1965, cited) credit Schade for how to make sinewave gratings on an oscilloscope but not for having made the first measurement.

> Included

Villadaite, Norcia et al., (cited) did not find spatial frequency specific effects at 8 cpd as in Pei et al,. (cited). So it's not really a replication of that study, but more of an extension to a measurement of CRF instead of SF tuning.

> Noted and altered

Page 12, better to state whether attention effects were consistent with response gain or contrast gain to make the result more concrete for the reader.

> Reynolds and Heeger (and Boynton) point out that the range of different types of change reported in the literature (response, contrast or mixtures of both) is explained by subtle differences in stimulus configuration (for example, the size and location of the 'probe' and any masking components relative to the size of the 'attentional spotlight'). We now note that the underlying model is one of contrast gain control (multiplicative amplification of the input rather than the output).

Discussion of time varying response paragraph on page 12 is not that relevant to contrast response function measurement, nor is the BCI section. These applications of SSVEP and could be deleted or the relationship to contrast response function made clearer.

> Originally, we envisaged this document as a primer on SSVEP contrast measurements for those coming to the field for the first time. In particular, we were struck by the large number of recent papers that use multi-frequency SSVEP contrast flicker as an input to BCI systems. There have been at least 400 such publications ('SSVEP AND BCI') in the last 5 years alone (compared to around 10 for amblyopia ('SSVEP AND amblyopia') and 80 for ('SSVEP AND clinical'). Some of these publications measure dynamic changes in covert spatial attention, some measure changes in featural attention,

some measure fixation. We know from personal experience that many of the researchers using contrast flicker as a BCI input would benefit from a slightly more applied guide such as this and the BCI section adds very little to the overall length of the manuscipt at this point.

Page, 14. It's not clear from the text how one would study feedback with the SSVEP, please elaborate.

In the paper we write that we would look for modulation of subcortical responses due to top-down signals. The most natural things to look for might be interocular normalization (because eye layers are segregated in the LGN) and attentional gain control (because covert spatial attentional modulation must arise in cortex). This is the type of feedback we are thinking of - not cortico-cortico feedback (which is hard to isolate with SSVEP).

In the last section, it would be worth mentioning that in order to "identify" a non-linear system, one needs multiple inputs. If the system is 2nd order, a minimum of two inputs is needed. If the system non-linearity is higher order than that, then more simultaneous inputs are needed (see Boyd, Tang and Chua 1983, IEEE Trans Circuits and Systems and Chua and Liao, 1991, Int. J. Circuit Theory and Applications).

This is a good point and we now include the Boyd reference in the section on fingerprinting.

Reviewer: 2

Comments to the Author

This manuscript is intended to be a review of the field of steady-state visual evoked potentials (ssVEPs) over the past 70 years. The authors state in the introductory paragraph "Here we will describe how an EEG method known as the steady state visually evoked potential (SSVEP) technique has contributed to our understanding of human contrast processing in health, disease and throughout development." The material included and the references cited, however, are quite limited and omit seminal work in this area.

We thank the reviewer for their helpful comments. As we noted above, our intention here was to provide an introduction to the SSVEP technique with appropriate simple code samples that would allow relative newcomers to the field to implement the technique. We are, however, keen to avoid misrepresenting the history of SSVEP research and are happy to include appropriate references.

1. All abbreviations should be defined upon first use.

Done

2. The following statement is unclear and inaccurate: "On-off flicker can drive independent populations of on- and off-cells in the retina once per cycle and can therefore produce a response at the fundamental flicker frequency, known as 1F, and its integer harmonics: 2F, 3F, 4F and so on. Counterphase flicker contains two transients per cycle and therefore does not produce a response at 1F, only at its even harmonics: 2F, 4F, 6F and so on." The term "on-off flicker" is typically used for simple luminance flicker without spatial pattern contrast. In the context of pattern stimulation, "on-off" can refer to "appearance-disappearance" of a spatial pattern with its contrast modulated in time. When the pattern appears and disappears into a uniform field of equal space-average luminance, the light and dark elements are modulated in counterphase. Any linear contribution to the response elicited by each set of elements (e.g., bars or checks) is expected to be canceled when summed at the electrode site on the scalp. Thus, the response at 1F to this stimulus is in fact a nonlinear response.

   Thank you for this explanation. We have now reworded the paper to make this clearer. We did take care to define what we mean by 'on-off' flicker in the preceding sentence but we now clarify the alternative 'appearence/disappearence' terminology. The point we were making was that odd harmonic responses (1F) are a feature of this type of modulation because the responses to the appearance and disappearance of a pattern (or mean field) need not be completely symmetrical (see Zemon, et al 1988 for an explicit demonstration).

This type of stimulation does not separate contributions from "independent populations of on- and off-cells." There are other ssVEP techniques developed that do tap ON and OFF pathways selectively, which are not cited here. Contrast-reversal stimuli which involve symmetrical modulation of light and dark elements cancel odd harmonic frequency components in the response, however, both kinds of stimulus techniques mentioned here elicit only nonlinear responses.

   Apologies that we have not been sufficiently clear here. We do not think we suggested any form of 'separation' and we agree that a) there are infrequently-used paradigms that can target ON and OFF pathways more effectively (for example, 'sawtooth' temporal modulations that provide asymmetrical onset or offset transients) and b) responses from spatial features of differing polarity will tend to be combined at the scalp as long as the spatial frequency is high enough. We have expanded this section to attempt to make the general point clear.

3. The statement "SSVEPs have proven to be an excellent measure of early chromatic processing as well" again fails to cite seminal work in this area.

   We have included three earlier papers by Regan on this topic.

4. The statement "This normalization, achieved through a computation called 'contrast gain control'" again neglects to cite the original work that defined

"contrast gain control" and the authors ignored a nonlinear biophysical model proposed to explain this phenomenon in ssVEP contrast response functions (which was also applied in developmental work on infants and children).

Contrast gain control is related to 'automatic gain control' - a electrical engineering concept that arose in the early 20th century. The concept of modulating the input (rather than the output) of the system became established in the 1930s and 40s (for example in WW2 radar systems). The specific idea of computing that modulation as a function of the local average signal power seems to have arisen around this time and was refined over the 50s and 60s (for example, in 'sidechain compression'). In the domain of vision, an early example of 'contrast gain control' (scaling the input - in this case luminance) is found in Sperling and Sondhi 1968.

The first formalization of the concept in the domain of spatial vision that included the critical concepts of an early, divisive gain control signal computed from some additional estimate of local contrast seems to appear in a seminal paper by Shapley and Victor, (1978). Application to contrast adaptation was demonstrated by Ohzawa, Sclar, Freeman, (1982, 1985) with a demonstration of instantaneous gain control in the human SSVEP by Bobak et al in 1988. A synthesis that included both adaptation and masking was proposed by Heeger 1992, while Foley 1994 showed an application to human psychophysical data. We have now cited these more foundational papers. We would be happy to include the relevant references to the non-linear biophysical model used in adult and developmental work if the reviewer can point us to the appropriate literature.

5. The description of Figure 3 and the nonlinear operations performed are unclear. They state in the text that Equation 2 ("hyperbolic ratio function") is used as a nonlinear transducer, but the figure legend states that a squaring operator was used for frequency doubling. A squaring operation would only produce a sinusoid of double the frequency of the input. The waveform shown in the left panel appears to be a full-wave rectified signal, which would contain theoretically an infinite number of higher harmonics. The text also states that "This results from the distortion of the input sine waves at high contrast due to a combination of the full-wave rectification and saturating non-linearity." When was full-wave rectification performed?

We agree that we went too fast here and the description of the signal processing was confusing. We now break out the two stages into (Figure 3) full wave rectification only (to model the contributions of on and off cells) - explicitly pointing out that this produces only even harmonics but does not alter the linearity of the CRF. We then show (Figure 4) the effect of passing this full wave rectified signal through a hyperbolic ratio function and illustrate the resulting saturation

of the CRF (Figure 4). We think that breaking the model out in this way makes the contribution of each part clearer and allows the reader to play with the code for separate stages. The input contrast is squared in Figure 4 simply because the exponent we choose in the saturating nonlinearity is '2' but this is a parameter that is often fitted rather than fixed and the reader is encouraged to experiment with the exponent to examine its effect on the resulting amplitude spectrum.

While there are, technically, infinite even harmonics in both Figure 3 and Figure 4, they decrease with frequency and we plot only three for reasons of space.

6. The authors use the term "frequency tagging" which has been used in the literature to refer to stimulating different regions of the stimulus field or fellow eyes with different sinusoidal frequencies and measuring the intermodulation (sum and difference) frequency responses that result from nonlinear interactions. They do not, however, cite the foundational work in this field.

As far as we are aware, the term refers to stimulating some *aspect* of the visual scene (a position or a feature) - not necessarily just a spatial region. There need not be more than one region (it is interesting, for example, to frequency tag a single feature or region and measure the effect of attending towards or away from it) and frequency tagged components can exist in the same location (for example, see the work on contrast normalization described above where the 'probe' and 'mask' have different tags and we explicitly examine the intermodulation terms). In fact much of this paper is concerned with discussing precisely these intermodulation terms and their origins. However, intermodulation is also not a 'necessary' part of a frequency tagging experiment: beautiful early work by Morgan et al (PNAS 1996) uses frequency tagging over space to examine attentional modulation purely in terms of the amplification or suppression of the F1 and F2 components. The tag need not even be visible - there is a fashion at the moment for 'rapid invisible frequency tagging' (RIFT) using flicker close to or beyond the CFF limit. In this sense, steady state paradigms are almost by definition frequency tagging paradigms because at least one feature or position is 'tagged'. In terms of using multiple input frequencies to characterise interactions between visual features, we think that Regan and Regan may have had the most comprehensive and some of the earliest descriptions of the method and we have credited them extensively throughout the paper. We are also happy to consider additional works if the reviewer can suggest them.

7. There is a large body of work that used ssVEPs to investigate contrast processing developmentally (infants through adulthood) and in various

disease states, which is mostly missed in this review article.

We touch on both these areas but including them in detail would require us to extend what is already becoming a very long paper.

8. Figures should be labeled.

It is possible that our text compilation stage failed to auto-generate the figure labels - we apologise if this is the case. The current version seems to be working correctly (in that each Figure has a label saying 'Figure 1', 'Figure 2' etc which is referenced in the main text).

Thus, the stated intention of this manuscript ("Here we describe how SSVEPs have been used to study visual contrast over the past 70 years") fails to be met. Only select articles from a few sources are covered, and therefore, the richness and value of this technique is lost.

The original document had well over 100 references and the current bibliography now contains 156 entries. Although we have tried to be fair, we have of necessity 'selected' from the literature and have doubtless omitted many important foundational works. We are happy to take suggestions if parts of the field have been misrepresented.

The original purpose of the document (now long eclipsed) was to serve as a convenient primer on the history of the technique and an interactive visualisation of key concepts implemented in Python and R using a novel 'executable paper' format. We think that getting this balance right means that we cannot include a comprehensive history of the entire field - in effect we must continue to be 'selective' and we hope that the relatively few papers we have cited here will serve as useful signposts to the richer literature.