Combining MRI and VEP imaging to isolate the temporal response of visual cortical areas

Thom Carney*[ac], Justin Ales[b], Stanley A. Klein[c]
[a] Neurometrics Institute, Oakland, CA
[b] Smith Kettlewell Eye Research Institute, San Francisco, CA
[c] School of Optometry, Univ. of California at Berkeley, Berkeley, CA

ABSTRACT

The human brain has well over 30 cortical areas devoted to visual processing. Classical neuro-anatomical as well as fMRI studies have demonstrated that early visual areas have a retinotopic organization whereby adjacent locations in visual space are represented in adjacent areas of cortex within a visual area. At the 2006 Electronic Imaging meeting we presented a method using sprite graphics to obtain high resolution retinotopic visual evoked potential responses using multi-focal m-sequence technology (mfVEP). We have used this method to record mfVEPs from up to 192 non overlapping checkerboard stimulus patches scaled such that each patch activates about 12 mm^2 of cortex in area V1 and even less in V2. This dense coverage enables us to incorporate cortical folding constraints, given by anatomical MRI and fMRI results from the same subject, to isolate the V1 and V2 temporal responses. Moreover, the method offers a simple means of validating the accuracy of the extracted V1 and V2 time functions by comparing the results between left and right hemispheres that have unique folding patterns and are processed independently. Previous VEP studies have been contradictory as to which area responds first to visual stimuli. This new method accurately separates the signals from the two areas and demonstrates that both respond with essentially the same latency. A new method is introduced which describes better ways to isolate cortical areas using an empirically determined forward model. The method includes a novel steady state mfVEP and complex SVD techniques. In addition, this evolving technology is put to use examining how stimulus attributes differentially impact the response in different cortical areas, in particular how fast nonlinear contrast processing occurs. This question is examined using both state triggered kernel estimation (STKE) and m-sequence "conditioned kernels". The analysis indicates different contrast gain control processes in areas V1 and V2. Finally we show that our m-sequence multi-focal stimuli have advantages for integrating EEG and MEG for improved dipole localization.

Keywords: multi-focal, m-sequence, VEP, visual cortex, fMRI, contrast processing, SVD

1. INTRODUCTION

Early visual processing can be studied with fMRI to demonstrate effects of attention and memory, but the slow hemodynamic response obscures interactions between cortical areas. EEG and MEG techniques provide much higher time resolution to potentially reveal the dynamic interplay between cortical areas, but the problem of isolating the signals from different visual areas has yet to be solved. Here, we present methods to profit from EEG/MEG time resolution, by providing a method to separate the responses of early visual areas. We first review limitations of traditional EEG methods, then introduce ways to separate the V1 from the V2 responses via cortical constraints given by MRI and fMRI measures.

Both EEG and MEG sum the response from multiple cortical areas. Separating nearby cortical sources such as in V1 and V2 is generally intractable, since arbitrary linear combinations of two close sources can typically produce the observed scalp voltage topography.[1] For this reason, simple dipole source modeling has not succeeded. To separate nearby sources, additional prior knowledge, such as from anatomy, is required. For example, distributed source methods constrain sources to the cortical surface given by MRI and estimate the topography from these sources. Distributed methods often have the drawback of producing sources that are excessively spread out. Weighted norm methods can produce more localized sources.[2,3,4] Dipole source modeling methods have recently started to incorporate cortical constraints to reduce the problem of nearby sources[5,6] to improve isolation of the dominant dipole, the likely V1

* neurometerics@speakeasy.net; phone 510-643-7571;

source. These studies compared fMRI localization to dipole source localization using EEG[5] or both MEG and EEG.[6] In both cases the focus was on the dominant dipole, which was assumed to be area V1. When EEG and MEG methods were combined, errors confirmed by fMRI, were on the order of 5-8 mm, a big improvement over earlier studies.

Success in separating sources with the help of prior knowledge depends on several factors, such as signal to noise ratio, proper forward model, and source orientation.[7,8,9] The problem is that sources can mix together, a phenomenon called the cross-talk,[10] or the rotation problem. We have previously examined rotations across individuals[11] and have shown how rotations can be minimized by taking Laplacians of the scalp topography.[12] However, using past methods, when sources are not sufficiently spatially separated the reconstructed source time functions are likely to be superposition of the true time functions. In this paper we show how an individual's own pattern of cortical convolutions can be used to solve the problem, a technique we call the cortical folding fingerprint.

With the advent of the multi-focal m-sequence evoked potential method we are beginning to see ways of generating detailed functional maps of early retinotopic areas of visual cortex.[13,14,15] The technique provides an electrophysiological means of estimating cortical magnification[16] and studying the M and P pathways in humans.[17] Using sprite graphics we have extended the standard method to increase the complexity of the typical multi-focal cortically scaled dartboard stimulus up to 192 checkerboard stimulus patches each independently contrast reversed according to an m-sequence. Cross correlation of the EEG signal with the m-sequence results in a series of response kernels for each electrode and each stimulus patch. Using this stimulus each patch activates about a $12mm^2$ area of V1. This extended multi-focal VEP technique in combination with the sparseness of cortical folding as a uniqueness constraint provides an effective method for localizing signal sources in early visual areas. The scalp voltage distribution strongly constrains the source magnitude (orientation). An abrupt change in source magnitude, as occurs for adjacent stimulus patches when the corresponding cortical patch of activation moves around a cortical fold, will result in a dramatic change in the scalp topography. The sparseness of cortical folding is important for separating source locations in different visual areas; a change in source magnitude in one area is not likely to coincide with a change in another visual area. Moreover, the multi-focal method seems to limit activation to early retinotopic visual areas where receptive fields are small in size. By using MRI/fMRI to constrain the sources for each stimulus patch to corresponding cortical surface locations in V1 and V2 we have been able to isolate each areas temporal response. Moreover, comparing the responses across hemispheres we have an independent verification lending confidence to the derived solution. Having shown that cortical folding is a critical component in isolating the response of nearby cortical areas, a technique is described that makes use of cortical folding without the need for costly fMRI/MRI recordings. The method involves more detailed visual stimulus manipulations and new statistical analysis to separate the time functions.

Now that we can isolate the V1 and V2 responses, we can examine the role of different areas in basic visual processes. For example, are rapid contrast gain control mechanisms found in only in area V1 or does it involve feedback from area V2? How fast are these processes? Using new m-sequence methods (STKE), we find gain control processes as fast as 32 ms and demonstrate different processing in areas V1 and V2 by using conditioned kernel methods.

2. CORTICALLY CONSTRAINED V1 AND V2 SOURCE IDENTIFICATION

This section describes our solution to the problem of identifying the temporal responses for early nearby visual areas, V1 and V2 in particular.
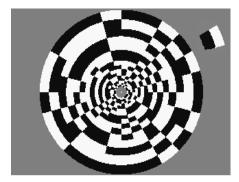

Figure 1 - One frame of 192 patch stimulus

2.1 EEG recording methods

Subjects viewed the display screen in a dark sound attenuating chamber. Stimuli were presented on a 60 Hz video display (1024 x 768 pixels) viewed from a distance of 109 cm (1.14 min square pixels). Luminance gamma correcting tables linearized the final display luminance. The stimuli were 15 deg. diameter dartboard shaped patterns with up to a 192 independent stimulus patches. The surrounding screen area and the central 0.5 deg diameter patch of the stimulus remained at the display mean luminance, 8.4 cd/m^2. A small bright fixation mark was placed at the center of the display which the subject fixated during each of the 1 min recording segments.

Dense 192 patch stimulus: Our primary purpose for using a stimulus with small stimulus patches is to enable us to identify the signal sources in early visual areas. With small patches correspondingly small areas of

retinotopically organized visual cortex (V1, V2 and V3) are activated. Adjacent stimulus patches activate adjacent cortical patches. The specific scalp voltage topography of the signal resulting from a particular patch of cortical activation is based on the local cortical surface normal. Given the extensive cortical folding, in moving from patch to patch in stimulus space the response topography should change according to the summed changes in the local cortical surface normals for early visual areas. The sequence of changes enables us to ultimately identify the source locations in the early visual areas.

The 192 sprite stimulus patch configuration (figure 1) consists of 8 concentric rings of 24 patches each. Each patch has two checks (one black and one white) in the radial direction and a single check in tangential direction. A single patch for the outer ring is shown in the upper right of figure 1. The size of each patch in the radial direction is scaled according to estimates of cortical magnification whereby the more eccentric the ring the larger the patch.[14] This scaling provides approximately equal areas of cortical activation irrespective of stimulus eccentricity. We estimate the cortical patch size in area V1 to be approximately 3.5 mm in the stimulus tangential and radial directions.[18] An example of one video frame of this stimulus is shown figure 1. The WinVis (W4M) psychophysical and physiological testing toolbox (www.neurometrics.com/winvis) was used to present and temporally modulate according to an m-sequence the 192 stimulus patches. Use of independent graphics sprites avoids the limitations of the LUT animation method of m-sequence stimulus temporal modulation. Details of the method are presented in the 2006 SPIE proceedings.[19] To study the response due to multiple stimuli simultaneously, each separate stimulus is driven with a temporally shifted sequence, resulting in a delayed impulse response for each stimulus when cross correlated with the originally, non-delayed, sequence. Multi-focal stimulation allows for an efficient way to combine many different stimuli in one experiment.[20] Each 16 bit m-sequence run was broken into 18, 1 minute recording segments with 1 second of segment overlap for combining segments during offline processing. The subject controlled the initiation of each recording segment allowing for brief rest periods between each segment.

Figure 2 - 96 electrodes

Ninety six active EEG recording electrodes were distributed over the scalp with a heavy weighting over the occipital pole as shown in figure 2. The EEG recording amplifiers were broad band with digital bandpass filtering applied offline.

2.2 MRI/fMRI recording methods

Magnetic resonance images were acquired at Stanford University using a 3-T GE Signa scanner. A special-purpose semi-cylindrical surface coil around the back of the head was used. Functional magnetic resonance images were oriented parallel to the calcarine sulcus. Eight functional images were acquired every 3 seconds using a two-shot, two-dimensional spiral gradient-recalled echo sequence with voxel size $2 \times 2 \times 3$ mm. Structural (T1-weighted) images were acquired in the same planes and with the same resolution as the functional images to co-register the functional and anatomical data.

The stimuli for the fMRI experiments consisted of rotating wedges and expanding annuli with a cycle of 72 seconds, resulting in five complete cycles during the 6-minute stimulus presentation. The wedge and ring were comprised of a flickering (reversal rate of 8 Hz) checkerboard.[21] Standard analysis tools were used for mapping the visual cortex. The three-dimensional cortex was unfolded onto a two-dimensional flat map for the region of interest, V1 and V2. White matter segmentation was performed to ensure a continuous gray matter surface for unfolding, using the FreeSurfer software package.[10,22] To isolate cortical areas V1 and V2 a Fourier transform was performed on the time series of the fMRI data for the rotating and expanding stimuli resulting in a response phase map corresponding to the locations in the visual field causing the excitation. The optimal hemodynamic delay was determined for each hemisphere for each subject by looking at the phase values of the rotating wedge runs. The chosen hemodynamic delay of 1.5 seconds resulted in the highest number of voxels excited by the correlated visual field and the lowest number of voxels excited in the non-correlated visual field. The Stanford mrVISTA tools were used to analyze and project the fMRI data onto the flat maps.[21,23] The result of this data analysis revealed the borders between visual areas and estimates of correspondence between visual field patch locations and sites of cortical activation (as shown in figure 3 a & b).

2.3 Our new algorithm for extracting V1 and V2 temporal responses.

The estimation of the time functions for multiple cortical areas from a single stimulus patch is severely under constrained. We have previously shown that by assuming a common time function for all cortical sources that correspond to a ring of stimulus patch within a visual area, an accurate estimation of the time function is possible (see

Appendix 3 for details.).[5,13] Here we apply additional cortical constraints given by the MRI/fMRI results. The voltages (V) recorded on electrodes (e) for a ring of stimulus patches (p), or hemi ring of patches for a single hemisphere is:

$$V(e,p,t) = \sum_{s=1}^{2} F_{true}(e,p,s)T_{true}(s,t) + noise(e, p, t). \qquad (2.1)$$
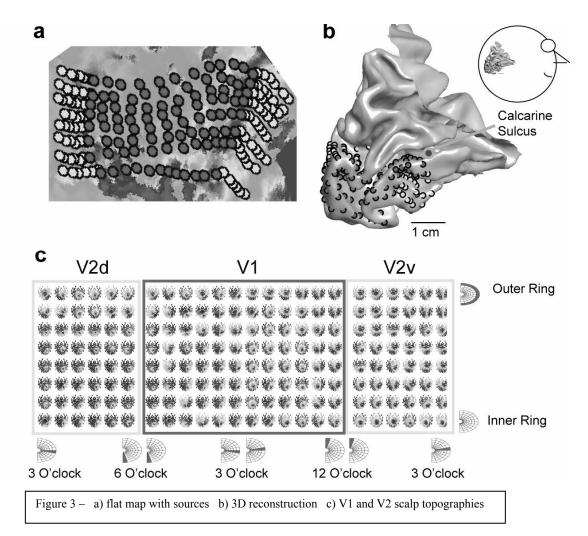
The index s refers to the early visual areas, here V1 and V2, with sources that contribute most strongly to the observed voltage V. F_{true} is the surface topography produced by each cortical source for a hemi ring of stimulus patches. Note that for the case of s_{max} =2 (areas V1 and V2, which are likely to account for about 70-80% of the total variance) T_{true} is a 2 by N matrix were N is about 150 time points. As discussed in Appendix 3, the time functions can be estimated by linear regression:

$$T(s,t) = \Sigma_{ep} pinv(F_{fmri}(s, ep)) V(ep', t) \qquad (2.2)$$

where $pinv(A) = inv(A^T*A)*A^T$ defines the pseudoinverse used to carry out the least squares linear regression.

The five parameters that specify the source dipole are given by the source location on the fMRI flap map and the orientation from the surface normal of the source location in the 3d MRI image. Since the fMRI based constraint along with using a 3 shell forward model for determining the surface topography is imprecise, the method for extracting cortical areas V1 and V2 time functions for each hemi-ring of stimulus patches requires an iterative process as outlined below.

Initial Conditions: The initial mapping of 192 stimulus patches to corresponding locations in areas V1 and V2 is derived from the fMRI data. The cortical source locations were estimated based on the flattened fMRI phase maps, known V1/V2 retinotopy and stimulus cortical scaling, which should achieve roughly equal source spacing in each visual area. This initial placement of source locations is shown in figure 3a which shows the flattened V1 and V2 phase map for one hemisphere. Once the V1/V2 boundaries are identified the sources are placed according to retinotopy with the inner stimulus hemi-ring of 12 patches placed near the foveal area of V1 and V2, which is the bottom row of sources indicated by small circles in figure 3a. The filled circles are the V1 source locations, while the open circles are the V2 source location estimates. Each row of sources moving up the figure corresponds to more eccentric hemi-rings of stimulus patches. For adjacent stimulus patches the V1 sources are separated by about 3mm. The border between open and filled circles in figure 3a corresponds to the upper and lower vertical meridians of the stimulus array. For each estimated source location a 3D MRI based reconstruction was used as in figure 3b, to extract the source dipole orientation for both areas V1 and V2. The sources in figure 3a are also represented on the 3D plot of figure 3b where the V1 sources (filled circles) are aligned around the T shaped calcarine sulcus. The V2 sources are located further from the calcarine as expected. For each V1 and V2 source location and orientation, the predicted scalp voltage topography is calculated using a 3 shell spherical forward model. Figure 3c shows the predicted voltage topographies after the source locations have been adjusted as described below. The box labeled V1 has 8 rows of 12 figures corresponding to the 96 contralateral stimulus patches that will evoke a response in this hemisphere. Each small figure is the predicted scalp voltage topography across the 96 electrodes, with the back of the head near the bottom of each small figure. The changing pattern of scalp topographies across the figures reflects the changes in the surface normal as the site of cortical activation changes from patch to patch along cortical folds. The boxes labeled V2d and V2v reflect the same effect of cortical folds in area V2 on the predicted scalp topographies. Since the areas V1 and V2 are presumed to be the primary contributors to the multi-focal VEP, the sum of the predicted V1 and V2 topographies times their respective time functions should approximate the recorded scalp voltages.

Figure 3 –   a) flat map with sources   b) 3D reconstruction   c) V1 and V2 scalp topographies

V1/V2 mapping optimization. After the initial source locations have been selected we iterate over the following three steps to find the solution that minimizes the error between the recorded and predicted scalp voltages.

Step 1. Do a linear regression to find the two source amplitudes at each time point (Eq. 2.2, T(s,t)) within a stimulus hemi ring of patches (12 patches). This step gives the best fit to the full dataset based on the dipole orientations specified by the MRI topography.

Step 2. For each patch location do an exhaustive search over the flat map for all points that are within about 3.5 mm of the original placement on the fMRI flat map for V1 and V2. For each source location on the flat map and its corresponding magnitude from the location's surface normal, use the 3 shell model to calculate the predicted topography across electrodes over time using the time functions of step1, $V_{pred}(e, p, t)$.

Step 3. Find the V1 and V2 pair of source locations for each patch that minimizes the sum of square error between the raw data and the predicted data.

$$SSE = \sum_{e,p,t}(V_{data}(e,p,t) - V_{pred}(e,p,t))^2 \tag{2.3}$$

Go to step 1 using the surface topographies ($F_{fmri}(e,p,s)$) from the new optimized dipole locations. Keep iterating until the SSE converges to a minimum. It should be noted that the source locations are never allowed to move more than 3.5 mm from the initial locations found from fMRI prior to step 1.

2.4 Dense cortical sampling – do small patches evoke a significant response

The 192 stimulus patches were cortically scaled to activate approximately equal areas of cortex, about a 3.5 mm square,[19] so unless the cortical sources for an individual stimulus patch are particularly badly located, each patch is expected to produce a similar response but with different topographies. Figure 4 shows the second order response kernel ([0 1] see Eq. 3.7) for a single patch as seen from the 96 electrodes. The electrodes near the bottom of the figure are positioned over the occipital pole while those at the top of the figure are on the forehead above the eye. Each waveform is the response for 500 msec following the patch reversal. As expected the strongest responses occur over the occipital area. While some patches may not evoke a response at a particular electrode, some subset of electrodes responds to each stimulus patch. This data set involved several days of recording to obtain a robust response and it is likely even smaller stimulus patches could be utilized.

2.5 Cortically constrained V1 and V2 time function estimation

From the initial conditions shown in figure 3 we iterated over the three steps described in section 2.3, which allows for up to a 3.5 mm shift at each source position to find the positions that the forward model for V1 and V2 best fit the actual multi-focal data across 96 electrodes and one hemi ring of 12 stimulus patches. The same procedure was independently applied to each of the 16 hemi rings of stimulus patches. In this way the common time function is applied to patches at the same eccentricity allowing for potential differences in response timing as a function of eccentricity. To validate the solutions we compare the V1 and V2 time functions for hemi-rings of opposite hemispheres but at the same eccentricity. Since the different hemispheres have unique cortical folding patterns this provides for a strong validation of the cortically constrained procedure. Figure 5 shows the 8 critical comparisons for V1 and V2. The upper time functions are for the outer stimulus rings while the bottom time functions are for the inner stimulus rings. The first thing to note is the similarly of time functions across hemispheres (solid vs. dashed line plots) at the same eccentricity. This provides an internal validation of the derived time functions. The differences that do occur are primarily at the inner ring of V1 and more so for V2. The fMRI map is poorer near the fovea so the cortical constraint is expected to be poorer for the inner ring. Area V2 is smaller so the signal is weaker and more subject to error which is also born out by the results. In summary, the data contributing to each hemi-ring are
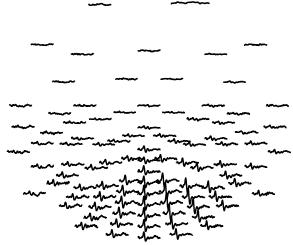

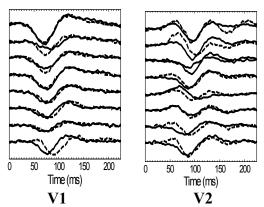Figure 4 – Response across electrodes to a single patch
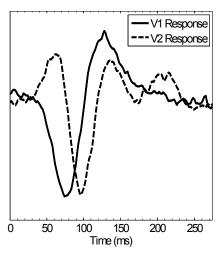

Figure 5 - Hemi-ring time functions

Figure 6 – V1 & V2

totally independent of each other, yet one sees consistency across hemispheres and across rings indicative of finding the actual time functions.

It might be argued that the original placement of the sources on the fMRI flat map was not an important constraint, so we shifted all the sources by 7mm on the flat map and repeated the search algorithm. This shift is small enough that the fMRI map still largely overlaps with the true retinotopic cortex, but since the dipoles are constrained to the orientation of the cortex it introduces a possibly large error in the source orientation. The consistency across hemi-rings (hemispheres) disappeared and the time function amplitudes were severely attenuated. The fMRI data was critical in limiting the time function solution space by constraining the source locations to the appropriate cortical patches (within about 3.5 mm).

No obvious consistent time function differences are evident across rings so to accurately estimate the actual V1 and V2 time functions we applied the same algorithm as before except the common time function constraint was extended across all the patches. Patches were grouped together across hemispheres and eccentricities. Figure 6 shows the V1 (solid line) and V2 (dashed line) time functions after yoking all patches with the same time function. In figure 6 two things are notable: 1) the V1 and V2 response appear to begin at the same time and 2) if anything the first peak of the V2 precedes that of the V1 response. There has been a controversy over what visual areas contribute to the early C1 component of the VEP with authors claiming either V1 or V2 is dominant[24] We presume the reason for this controversy is the rotation problem described in our 2006 SPIE paper.[19]

Most, if not all, prior examples of source identification were not pure V1 or V2 sources but some linear mixture of the two sources. Both the latency and identity of the sources can be misestimated, while still obeying general rules as to the assumed shape of the calcarine fissure and its impact on the striate source orientation. A source orientation flip near horizontal meridian is often the primary basis for deciding if a source is striate or prestriate. However a source with a significant mixture of V1 and V2 signals will exhibit the orientation flip. We contend that in most cases the C1 component of the VEP is some mixture of V1 and V2 sources. While the anatomy might indicate that V1 precedes V2 in signal processing,[25] different conduction velocities of cell classes and alternative pathways could contribute to an early V2 response. Moreover, depth electrodes in macaque V1 and V2 show a similar ERP profile to the responses in figure 6, with V1 and V2 waveforms having opposite polarities and similar onset timings.[26] It has not escaped our notice that the main V1/V2 activity has a coupled oscillator nature[27] such that the rate of change of V2 activity is roughly facilitated by V1 activity and the rate of change of V1 activity is roughly inhibited by V2 activity. This relationship isn't precise since as will be shown in our Gabor fit to the time functions (see Section 4.1, fig. 9) the V1 and V2 time functions are about 7 and 10 Hz near the peak activity.

At the 2006 SPIE meeting we presented results of a principal component analysis (SVD) of this same data set.[19] As it turns out the first principal component is similar to the V1 solution based on this new cortically constrained method. Zhang and Hood using a multi-focal stimulus with just a few electrodes have reported that the first component of SVD is a "relatively pure V1 component and can be used to study the processing in V1".[28] Their conclusion was partly based on the component polarity changing as a function of patch location in accordance with expected average V1 cortical surface conformation. While this is consistent with a V1 source, it also true that a significant mixing of V1 and V2 components could still result in source magnitudes that might be expected from sources along the upper and lower bank of the calcarine sulcus. Moreover, when we recorded from 8 subjects using similar multi-focal stimuli we found substantial individual differences in the first principal component which could involve a rotation between principal components. The individual differences were greatly reduced when components were rotated to a common subspace (see Dandekar et al, figure 12a-d).[11] Even small errors in estimating the very dominant V1 time function can have major impact on isolation of later areas that are much weaker. Finally it must be mentioned that the second principal component was different from the cortically constrained V2 time function. This is expected of the orthogonality constraint of SVD and the rotation problem.

3. COMBINING EEG AND MEG FOR SOURCE IDENTIFICATION

We have evaluated dipole localization accuracy based on fMRI results using similar stimuli and found the dipole solutions were about 1 - 1.5 cm from the fMRI expectation.[5] More recently Sharon and colleagues compared source identification using EEG and MEG separately as well as combined.[6] Localization perfomance based on EEG data was similar to our earlier findings. However, when EEG and MEG are combined the source localization was substantially improved over either method alone. Huang, et al. carried out simulations for testing a new approach for integrating EEG and MEG.[29] They used MEG data to determine the optimal conductivity model for the EEG analysis and then used

EEG to estimate the radial sources which MEG does not see. Other methods have been used to combine EEG and MEG data by scaling the EEG signal based on measurments made when only a tangential source was present.[30] The problem is the time functions for EEG and MEG appear different. There are at least two possibilities to account for the difference in EEG and MEG time functions:

1) It could be possible that even though the EEG and MEG dipole current sources have the same location and orientation, there might be something about the physics of pyramidal cell currents that allows different sources for the two types of measurements. Note that the difference being proposed here is not simply independent radial and tangential currents (what has been called a 'rotating dipole'). A standard rotating dipole would have it rotating for the EEG compents as well as for the EEG/MEG different components.

2) It could be that the different time functions measured by EEG and MEG is because multiple areas are responding to the stimulus and the dipole fitting methods are not truly isolating the individual sources. For example, the measurement of the V2 dipole is getting different amounts of intrusion from other nearby visual areas depending on whether the EEG or the MEG forward model is being used, resulting in an apparent difference in time functions.

It is our opinion that there is no solid basis for the first possibility. For that reason it has been our strategy to minimize the number of areas contributing to the responses. In our multifocal stimulation we use dartboards with many relatively small patches and the stimuli we have been using are m-sequence binary noise stimuli being updated at 60 Hz. We have found that principal component analysis shows that when using these stimuli between 70-85% of the variance of the response can be attributed to just two principal components.[11,15,19] We expect that since there are just two dominant components the time functions of EEG and MEG might be similar since intrusion by additional areas would be minimized. In this section we present data validating this assumption.

Experiments using simultaneous MEG and EEG recording were carried out on a different subject at the University of Montreal's CTF instrument with 271 MEG and 57 EEG channels. The stimuli were sampled at 60 Hz and the responses were sampled at 480 Hz in a time locked manner. Here we report our analysis of the results of totally separate SVD analysis of the EEG and MEG data. SVD gives:

$$V_M(e, p, t) = \sum_{c=1}^{N} S_M(c) E_M(e, p, c) T_M(c, t) \tag{3.1}$$

$$V_E(e, p, t) = \sum_{c=1}^{N} S_E(c) E_E(e, p, c) T_E(c, t) \tag{3.2}$$

where N equals the number of time points (because they are fewer than the number of patches times number of electrodes), the subscripts M and E correspond to the MEG and EEG data, and the strength of each component is given by $S_M(c)$ and $S_E(c)$. The functions E and T that are output by SVD are orthonormal. That is:

$$\sum_{t} T(c, t) T(c', t) = \delta(c, c') \tag{3.3}$$

$$\sum_{e,p} E(e, p, c) E(e, p, c') = \delta(c, c') \tag{3.4}$$

where $\delta(c, c')$, the Kronecker delta function is unity for c=c' and zero otherwise. The summations are over all the time points for Eq. 3.3 and over all patches and electrodes for Eq. 3.4. The square of S(c) is the contribution of the each component to the total sum of squares (TSS) of the original data, where:

$$TSS = \sum_{e,p} S^2(c) \tag{3.5}$$

It is useful to define the fraction of TSS contributed by each component:

$$SS(c) = S(c)^2/TSS \tag{3.6}$$

One of the impressive aspects of our m-sequence multifocal stimulation is that very few components contribute to the total SS, as shown in the top pair of plots of figure 7 for subject 1 with the MEG and EEG results shown on the left and right, respectively. For this subject, the fractional contribution to TSS from first three components of the MEG data are: SS(c) = 65%, 18% and 3%. For the EEG data the contributions are: 48%, 14% and 6%. The large amount of variance accounted for by the first few components is impressive, especially for the MEG data.

The time functions for the first two components (C1 and C2) are shown in the middle pair of panels. The time functions that we used for the SVD are a concatanation of four independent time functions that are obtainable from our non-linear m-sequence analysis. The first three are the first three cuts of the second order kernel and the fourth is the first cut of the fourth order kernel. The second order kernels, specified by the notation [0 k] are defined by:

$$[0\ k] = \Sigma_t\ y(t)\ x(t - \tau)\ x(t - \tau - k) \qquad (3.7)$$
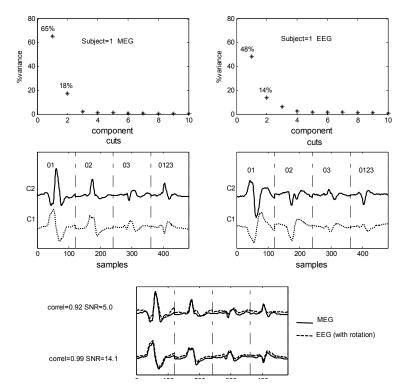
The fourth order kernel, is specified by:

Figure 7 – Equivalence of EEG and MEG time functions

$$[0\ 1\ 2\ 3] = \Sigma_t\ y(t)\ x(t - \tau)\ x(t - \tau - 1)\ x(t - \tau - 2)\ x(t - \tau - 3) \qquad (3.8)$$

The various kernels are intimately related to the rapid contrast gain control of the system as discussed in Section 4.1.

The middle panels of Fig. 7 compare the EEG to the MEG. Notice that the EEG and MEG time functions are inverted from each other. That is not surprising since both T and E in Eq. 3.2 can have a sign change with the resulting data being unchanged. Much more interesting is the double bump in the [0 2] kernel of the second component of the EEG that is not present in the matching time function of the MEG. One possibility is that there is a true difference in the EEG and MEG waveforms. Another possibility is that the EEG components are rotated as compared to the MEG. Since the leading two components of the MEG account for 83% of the variance we will take the leading two MEG components as defining the subspace of the true sources. Since the EEG topography and forward model is very different from that of the MEG, there could easily be a rotation between the two sets of time functions.

In the bottom panel that plots the EEG (dotted line) with the MEG (solid line) we allowed a one parameter rotation to align the two functions. The EEG has been shifted vertically slightly to make it visible. It is seen that the two functions across the 500 samples are well matched. It is noteworthy that after the rotation the double bump in the [0 2] cut of the EEG 2nd component has been greatly reduced after the rotation. Since the first component is substantially stronger than the second one, it is not surprising that the peak of the first component has 'leaked' into the weaker second component. After the rotation the correlations are 0.990 for the first component and 0.923 for the second component. Alternatively, one can express the similarity of the two functions as an rms signal to noise (SNR) as given by:

$$SNR = \sqrt{\{sum(T_M + T_E)^2\ /\ sum(T_M - T_E)^2\}} \qquad (3.9)$$

where the summation is over the 500 time samples. The SNR is 14.1 and 5.0 for the first two SVD components.

We will call this subspace the "V1/V2" subspace. We are well aware that it is too early to call it a V1/V2 subspace, which is why we put it in quotes.

The bottom panel of Fig. 7 fulfills our Section 3 goal of demonstrating that our multi-focal m-sequence stimuli do an excellent job of restricting the main EEG and MEG responses to just two or three components (likely visual areas). With the m-sequence stimuli the EEG and MEG time functions are likely to be very similar.

4. CONTRAST PROCESSING: new methods and differences between visual areas.

Using multi-focal m-sequence methods we have been able to extract the temporal response of early visual areas. We now have tools to explore how contrast sensitivity, contrast masking and gain control processes occur within and between visual areas. Several groups have used multi-focal stimuli to characterize the system response as a function of patch contrast.[15,17] In particular, Hood et al. using multi-focal m-sequence stimuli have compared response strength as a function of contrast with an established model of single cell responses in area V1. The VEP response followed the model well up to a midrange of contrasts but saturated soon after, a marked departure from the single cell model as well as other models derived from psychophysics.[31,32]

Moreover, the response latency didn't change with contrast in accordance with the single cell models. Groups using fMRI sometimes find striate and prestriate saturation at around 30% contrast while others find a more linear response profile. Recently one group[33] has used MEG recording to study the contrast response to a pattern onset stimulus and report linear response function in striate and a saturating response in prestriate cortex. While the findings thus far are contradictory, it does seem that the shape of the response function may vary across visual areas. Contrast gain changes dependent on stimulation rate could also be a factor in the different results.[34]

4.1 Conditioned Kernels and Rapid Contrast Gain Control: Comparing MRI based sources to SVD based components.

Section 3 compared the MEG and EEG time functions. As was discussed in that section the time functions being compared were concatenations of various cuts of nonlinear Volterra kernels. These Volterra kernels are calculated (Eqs. 3.7 and 3.8) as cross products of the response with product of shifted stimuli. Consider, for example the second cut of the second order kernel (Eq. 3.7 for k=2)

$$[0\ 2] \qquad h_2(\tau, \tau\text{-}2) = \Sigma_t\, y(t)\, x(t - \tau)\, x(t - \tau - 2) \qquad\qquad (4.1)$$

If the stimuli were sparse flashes this kernel would be the response of the system to a pair of flashes two video frames apart (assuming the stimulus, x, is sampled in frames) minus the responses to the two flashes presented separately. The time index t and τ are sampled at the faster response sampling rate with t-τ constrained to be integers in the summation. Our stimuli are not sparse flashes, but are rapid random reversals. Thus [0 2] is equivalent to time locked addition of the response to all occurrences of the same state occurring at time t and t+2, and subtracting the response to all occurrences when the stimulus has opposite sign. This amounts to adding responses with an even number of reversals between the two time points and subtracting the response to an odd number of reversals. This can be written as:

$$[0\ 2] = \{?\ ?\ 0\ 0\ ?\ \} + \{?\ ?\ R\ R?\ \} - \{?\ ?\ 0\ R\ ?\ \} - \{?\ ?\ R\ 0\ ?\ \} \qquad\qquad (4.2)$$

where 0 and R mean no reversal and reversal, and the question mark means we don't care whether there is a reversal or not at that frame. Note that the square bracket represents the Volterra kernel and the squiggly bracket is the reversal kernel. Alternatively, we could have removed the first and last question mark:

$$[0\ 2] = \{?\ 0\ 0\} + \{?\ R\ R\} - \{?\ 0\ R\} - \{?\ R\ 0\} \qquad\qquad (4.3)$$

Or we could make the question mark in Eq. 4.3 explicit:

$$[0\ 2] = \{0\ 0\ 0\} + \{R\ 0\ 0\} + \{0\ R\ R\} + \{R\ R\ R\} - \{0\ 0\ R\} - \{R\ 0\ R\} - \{0\ R\ 0\} - \{R\ R\ 0\} \qquad\qquad (4.4)$$

A similar analysis could be done for [0 1], [0 3] and [0 1 2 3] which together with Eq. 4.4 comprise the four kernels that were shown in Fig. 7. The four equations can be inverted to give the four conditioned reversal kernels:

$$\{0\ 0\ |\ R\} = \{0\ 0\ R\} - \{0\ 0\ 0\} = [01] + [02] + [03] + [0123] \qquad\qquad (4.5a)$$
$$\{R\ 0\ |\ R\} = \{R\ 0\ R\} - \{R\ 0\ 0\} = [01] + [02]\ - [03] - [0123] \qquad\qquad (4.5b)$$

$$\{0 \, R \mid R\} = \{0 \, R \, R\} - \{0 \, R \, 0\} = [01] - [02] - [03] + [0123] \qquad (4.5c)$$
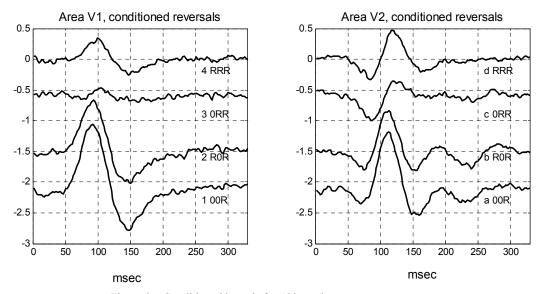$$\{R \, R \mid R\} = \{R \, R \, R\} - \{R \, R \, 0\} = [01] - [02] + [03] - [0123] \qquad (4.5d)$$

Figure 8 – Conditioned kernels for 192 patch

The conditioned kernel notation $\{0 \, R \mid R\}$ means one is looking at the response to a reversal that is conditioned by a preceding reversal that in turn is preceded by a no reversal. A general formalism for conditioned kernels has been developed by Klein[35] and applied to ganglion cell data where the conditioning goes back 90 msec for both center and surround. Subtle shifts on the order of 1 msec in spike firing time can be attributed to the stimulus condition 75 msec in the past.

Figure 8 shows the conditioned reversal responses for the 192 patch data discussed in Sections 2 and 3 above. We previously only displayed the first cut of the second order kernel ([0 1]). The [0 2] and [0 3] Volterra were calculated at the same time as the previously reported [0 1] kernel. We unfortunately didn't calculate the [0 1 2 3] kernel that would have allowed for the full expansion, but based on other data, the [0 1 2 3] kernel (as well as [0 3]) is smaller than [0 1] and [0 2]. Not surprisingly, the four conditioned kernels are all somewhat similar to the [0 1] kernel that is the strongest. It is clear that the strongest response occurs for $\{0 \, 0 \mid R\}$ when the visual area has had a minimum of two frames of rest. When the primary reversal is preceded by only a single frame of rest $\{R \, 0 \mid R\}$ the response is decreased slightly. When the primary reversal is preceded by a reversal the response behavior differs for V1 and V2 and that difference depends on whether the preceding reversal is preceded by another reversal. In order to clarify these dependences we have carried out an 11 parameter Gabor fit to the four curves of each panel of Fig. 8. .
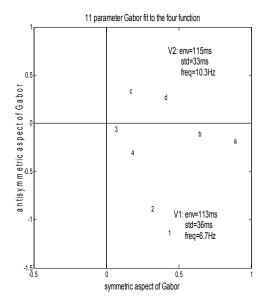

Figure 9 – Gabor fit to condition kernels

$$G_i(t) = \exp(-t - t_0)^2/2\sigma^2) \; (a_i \cos(2*\pi*(t-t_0)) + b_i \sin(2\pi \, f \, (t-t_0))) \qquad (4.6)$$

Three of the parameter, the time (msec) of the envelope peak, t_0, the standard deviation (msec) of the Gaussian envelope, σ and the temporal frequency, f (Hz) were determined using Matlab's lsqnonlin, linear regression routine. Inside the function being searched the eight coefficients a_i, and b_i of the symmetric and anti-symmetric parts of the Gabor were determined by linear regression. The 11 parameters are presented in Fig. 9 for both V1 (numbers 1-4) and V2 (letters a-d). The symmetric and anti-symmetric aspects of the Gabor function are presented on the x and y axes respectively. It is seen that the strong responses of V1 (1 & 2) and V2 (a & b) lie near the negative vertical and positive horizontal axes respectively.

One of the most remarkable aspects of the conditioned kernel response is that the V1 response to {0 R | R} is nearly extinguished. This can be thought of as the gain control of the gain control or in other words the preceding reversal is extra strong since that prior reversal in turn is preceded by at least one blank period.[36,37] That extra-strong reversal suppresses the primary reversal. For V2 the suppression is not nearly as strong.

An interesting question is whether the same pattern occurs in other observers. Fig. 10 shows the four Volterra responses for both EEG and MEG for a new observer. When combined according to Eq. 4.5 the conditioned reversal responses are obtained, and plotted in Fig. 10. There is a striking similarity of these responses, obtained with SVD and the responses for the previous observer that was obtained with the far more complicated MRI/fMRI based dipole source localization. One striking difference in these plots is that each curve is doubled. The doubling is due to our having interleaved two identical m-sequence stimuli when carrying out the experiment. Each stimulus segment lasted 6 sec. The observer's task was to attend to a dim flash that was on either the right or the left side.
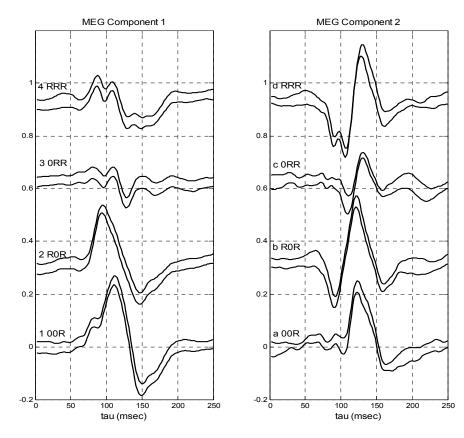

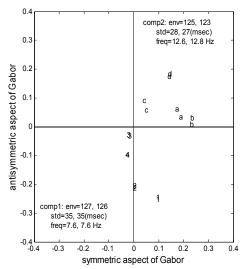Figure 10 – MEG condition kernels (test-retest reliability)

Figure 11 – Gabor fit conditioned kernels

Although we didn't find an attention effect (possibly the very strong flickering stimuli attracted exogenous attention across the full visual field) we had the side benefit of collecting totally independent sets of data that were gathered on

exceptionally well-matched conditions. The doubling is also seen in the impressive test-retest agreement in Gabor phasor plots of Fig. 11.

The other difference from Fig. 8 is that rather than calling the right and left panels V1 and V2, they are called component 1 and component 2, since they come from the SVD analyses. We chose to show the conditioned kernels for the MEG rather than the EEG since although they looked similar, the first two MEG components accounted for 83% of the variance, whereas the two EEG components accounted for 62% of the total variance.
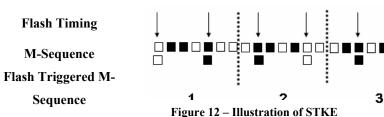
The pattern shown for these MEG conditioned kernels has a striking similarity to the EEG kernels shown in Figs. 8 and 9. Again the {0 R | R} kernel for component 1 is nearly zero. Although the general structure of the plots for the two observers is similar, there are also differences. We must remember that the components 1 and 2 are not to be trusted as being identified with V1 and V2. The components are forced to be orthogonal by the SVD algorithm, and we expect some of that orthogonality is produced by the V2 source "borrowing" some of the strong V1 component. This leakage of V1 into the V2 source leaking might account for the shared undulation at $\tau = 100$ msec. We anticipate that if the proper MRI/fMRI based dipole rotation had been made, the phasor plots of Figs. 9 and 11 would be even more similar.

4.2 Rapid and intermediate speed contrast gain control - STKE

We have been working with an m-sequence based system we call state triggered kernel estimation (STKE) to address the specific question, how does the system impulse response change at different time delays after a sudden contrast change? The sudden contrast change violates the stationarity assumption and would have been impossible to analyze with most previous white noise methods. The system dynamics were analyzed by implementing an extension to the m-sequence stimulation method that allowed us to trigger the kernel estimation by a change in the stimulus.

Menz et al. developed a clever method using m-sequences that allows one to track the dynamic changes in evoked responses that occur in response to a state change,[38] such as occurs during light adaptation. The STKE method is illustrated in figure 12 for a toy example. The line of the figure shows 4 repeats of a standard 3 bit m-sequence that consists of 7 frames per sequence. The open and filled symbols indicate the stimulus state given by the binary m-sequence. What is unique about this method is the addition

Flash Timing

M-Sequence

Flash Triggered M-Sequence

Figure 12 – Illustration of STKE

of a much slower stimulus whose repetition time is a power of 2 number of frames (2^n frames), in this example, every four frames (shown by the arrows). This slow stimulus was a high contrast pattern for producing masking or gain control. Because the distance between flashes is 4 frames we need 4 repeats of the m-sequence to be able to reconstitute a full m-sequence for different time lags. For example, to estimate response kernels for the time coincident with the flash you would take the 7 responses from the same time as the flashes. This is shown in the bottom row. If you examine the sequence in the bottom you will see that it is a complete copy of the original repeated m-sequence. The special thing is that each one of these data points comes from a time when a flash is presented simultaneously. Also one could choose the responses one frame after the flash to make a perfect m-sequence that follows the slow flash by one frame. With more realistic (much longer) sequences one is able to have multiple patches and multiple locations for the

Figure 13 - The nomenclature used in presenting the results

cue or masking stimuli. Since this method is based upon an m-sequence stimulus it is straightforward to extend the method to include having both flashes and probes for more than one stimulus location simultaneously. Since the multi-focal method works by presenting a temporally shifted version of stimulus at each desired location it is possible to independently stimulate multiple locations.

The main feature of this method is that it is focused on finding the changes in kernels as a function of offset from another event. This event in this case is a large change in stimulus contrast. Of interest is the changes in the responses to probes presented at different times following the state contrast change. Time differences such as these have been referred to as stimulus onset asynchrony (SOA). Figure 13 contains a schematic of the stimulus in order to illustrate the timing concepts and our terminology. Two different time indices are used in this paper: delta and tau. These indices are used to differentiate between what is used as the reference time. Delta is the time from the state transition to the stimulus probe and tau is the time from the probe to the time of response. T, the time from the state
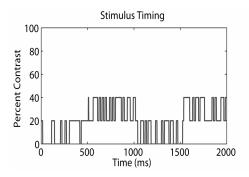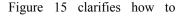

Figure 14 – M-sequence contrast modulation

transition to the response is T, which equals delta plus tau. Delta and tau can have much different sampling rates. In our case delta is sampled at the 60 Hz display rate and tau is sampled at the EEG response sampling of 512 Hz.

To measure the impulse response at intervals after a contrast step a standard checkerboard 60 Hz pattern appearance m-sequence probe stimulus was presented to which a masking pedestal as added that alternated between a mean luminance field and a high contrast pattern at 1 Hz. The state triggered kernel estimation (STKE) method was used to show how the introduction of a contrast pedestal dynamically changes the response of the visual system.

Four male volunteer subjects participated in the study. Electroencephalograms were recorded while wearing a cap with 64 active electrodes based on the standard 10-20 configuration. EEG data was digitally filtered with a pass band of 2-100 Hz. Each run was divided into one minute recording periods, each separated by a subject defined rest interval. The stimulus was defined on an annulus with an inner radius of 1 degree and an outer radius of 8.5 degrees. The pattern within this annulus was a checkerboard with 48 spokes and 16 rings (similar to figure 1). The contrast of this pattern was modulated according to an m-sequence. Figure 14 illustrates the temporal sequence of both the slow pedestal and

the fast test probe. There are two types of modulation going on simultaneously, one is periodic (the "state") the other is random. The timing of the random stimulus is controlled by a 63 frame m-sequence that specified either 0 or 20 (0 or 60 not shown) percent contrast. While this random m-sequence was ongoing a contrast pedestal was modulated by a square wave at approximately 1 Hz, (64 video frames/cycle at 60 Hz). The contrast pedestal was either 60 or 0 percent. At any instant the contrast was 0, 20, 60 or 80 percent. We then extracted the response corresponding to the pattern appearance by using a variation of the fast Walsh transform.[39] The method allowed us to extract responses time locked to many different time points during the changing pedestal contrast.
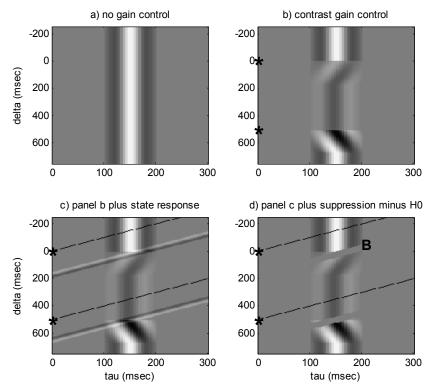
Figure 15 clarifies how to

Figure 15 – Simulated STKE response profiles

interpret the format in which STKE data is presented. The response to a stimulus x(t) can be written as:

$$V(t) = \Sigma_\tau \, x_p(t - \tau) \, h_p(\Delta, \tau) \; + \Sigma_T \, x_s(t - T) \, h_s(T) \tag{4.7}$$

The first term is the response to a probe that occurs at time t-τ. The magnitude and timing of the response can depend on the time Δ between the time of the state transition and the probe, which is why h_p depends on both Δ and τ. The second term is the response, specified by $h_s(T)$ to the state transition that occurs at time t-T. Eq. 4.7 is a mathematical way of expressing what is shown in Fig. 13, but with a response from the state transition as well as the probe.

Figure 15 is a simulation of the response of a system using our standard STKE plotting format. The upper left panel (15 a) shows V(τ) the response to the probe as a function of τ. In this made-up example the system starts responding 100 msec after the probe and reaches a positive peak 150 msec after the probe. Panel 15b shows how the response is altered if at t = zero, the background contrast is strongly elevated and at t=500 msec it is turned off. The pair of asterisks on the ordinate indicates the time of the state transition. In this simulation the response amplitude to the probe is reduced instantaneously to probes following the transition. The latency of the peak response is also decreased (a faster response to higher contrast) but this timing shift gradually get stronger as the gain control of successive probes adjust to the new state contrast. Panel 15c adds the response to the state change to the STKE plot. The state change response is taken to be a quick one cycle response at about 150 msec after the state change. Points with a fixed state change response are show to have a unity slope on the STKE plot. This counterintuitive results is because the time after the state change, T, is given by T = Δ + τ, as can be seen from Fig. 13. An especially important line is the dashed line that starts at the asterisk and has a slope of unity (if the axes were equally scaled). The dashed line is the line of constant T=0, where any effect on the response due to the state transition that occurs before the slanted T=0 line would be a violation of causality. Referring back to Fig. 13 could be helpful in clarifying why this is so. Panel 15d, is the same as panel c except that the linear portion of the state change response has been subtracted off, as it is in our plots of real data to be shown in figure 16. Although the linear part is subtracted off, possible nonlinear effects remain and in this case we assume that the state response suppresses the probe response, not unreasonable because the state change is typically much stronger than the probe response. For large values of τ, this suppression can come before Δ=0. This is an example of backward masking, as shown by the letter B on panel d. The apparent violation of causality is illusory since all it means is that the probe comes before the state change and the state change response is fast enough to suppress the probe response at a positive time T after both probe and state change. Our data show examples of this backward masking as can be seen by looking at the region in the real data corresponding to point B in the simulations.

4.2 Rapid and intermediate contrast gain control revealed using the STKE method

The recorded EEG data was run through an independent component analysis (ICA) to denoise the data. The ICA revealed a strong component common across the 4 subjects. We used this component to extract time courses from each subject, these time courses were then averaged across subjects. The averaged evoked responses of this component are displayed in figure 16 for the 60% contrast pedestal stimulus. The vertical axis is ordered with increasing deltas going down as indicated on the right hand side of the figure. The horizontal axis has increasing taus going towards the right. Since the stimulus is periodic the definition of the first delta is arbitrary. Therefore, the plot has been doubled in the vertical dimension to both emphasize the periodic nature of the stimulus and facilitate comparisons of all deltas.

The contrast pedestal has a dramatic impact on the evoked potential. This change is not just a modulation of the response amplitude, but also change in the shape of the signal. For example, the initial response polarity changes with the contrast step change as seen near the arrow in figure 16. Changing of the signal phase/frequency as a function contrast pedestal has been shown in the steady state VEP before.[40] But with the steady state VEP it is impossible to measure the speed of these changes, while in the STKE plot it can readily be seen that the response changes rapidly with the application of a contrast pedestal. It is so rapid that the very first video frame with or without the pedestal has a different response. Recovery of the response after the contrast pedestal is turned off can also be seen at deltas 533-1067 and repeated at 1600-2133.

One of the most interesting aspects of the data is the changes that occur in the 50 milliseconds around the pedestal change, the brief diagonal pattern. This effect is pointed out by the surrounding white oval near the bottom of figure 16. That edge of the figure is aligned with the transition from 0 to 60% contrast pedestal. At first glance this diagonal pattern may seem to violate causality, because the changes appear to precede the pedestal change but this is consistent with a slower gain control system causing backwards masking. This is because real time is the sum of delta and tau (see figure 13). This backward masking was detailed earlier in reference to the inset **B** in figure 15d.

Besides the relatively rapid contrast gain effects, figure 16 also captures a slower recovery process which is pointed out by the upright oval in the figure. In this case after the pedestal is turned off at 533 msec the response strength slowly increases over about the next 300-400 msec.
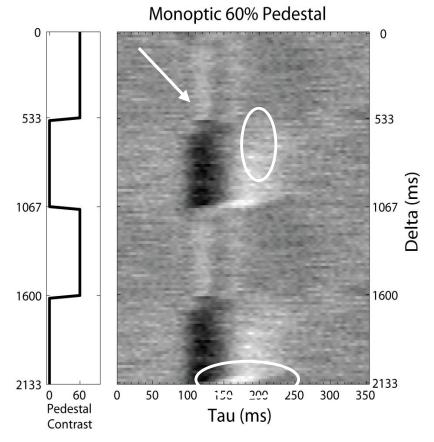

Figure 16 – STKE contrast gain control

5. DISCUSSION

5.1 V1 and V2 temporal response functions using cortical constraints.

The standard multi-focal m-sequence method is a powerful tool to study cortical areas that are retinotopically organized. Independent responses to 192 stimulus patches, each of which activates about a $12mm^2$ area of cortical V1, provides a means to very effectively incorporate cortical constraints in finding the time functions for areas V1 and V2. Robust cortical constraints are necessary for disambiguating the nearby sources of activity in early visual cortex where inaccuracies in head model and other types of noise can make the scalp topography from many different nearby sources appear the same. The problem is under constrained which is why functional anatomy given by MRI and fMRI can dramatically constrain the solution space. In the past the primary anatomical constraint was based on finding a source component that has inverted polarities for stimuli in the upper versus lower hemi-fields. Anatomically, V1 is identified by the calcarine sulcus, with a retinotopy that predicts a flip in the orientation of a dipolar source as it moves around the sulcus. This gives a good prior on how a V1 source should behave, however nearby sources could still intrude causing a change in the source orientation while still conforming to this model. Our new method uses not just a general cruciform model for the shape of the calcarine sulcus, but rather each individual subject's unique folding fingerprint to identify the locus of the VEP components. The V1 and V2 VEP components were found to begin at approximately the same time with opposite polarities and different peak latencies. In light of the component rotation problem, the controversy over what visual areas contribute to the early C1 component of the VEP is not surprising. This is why it is important to have a method with an internal consistency check to verify the purity of source reconstruction. The similarity of source time functions across hemispheres provides necessary source validations.

The method presented in this paper provides the ability to separate the V1 and V2 source contributions from the mfVEP waveform. The method not only provides the responses specific to the areas in question, but also contains the ability to internally verify the consistency of the extracted waveforms. The latencies for the V1 and V2 temporal functions are nearly identical yet the overall waveforms are quite different which probably reflects the dynamics between visual areas. The same general procedure could be extended to identify the source time functions for later retinotopic visual areas. Understanding the nonlinear neural dynamics of the multiple early visual areas could be an important step in understanding more general interplay of closely connected brain areas.

5.2 Contrast processing in early visual areas

Modeling gain control processes has been a very active area in visual psychophysics. Much of the purported physiological substrates for the mechanisms comes from vertebrate single cell recordings. Recent MEG, EEG and fMRI data are inconsistent as to the differential contributions of visual areas. We present two new methods of examining contrast gain control in humans using EEG and MEG. It appears that rapid gain control is different across visual areas which offer a new challenge to modeling human contrast processes.

Conditioned kernels reveal rapid contrast gain processes in different visual areas: M-sequence conditioned kernels were used to compare the response to a pattern reversal when the reversal was preceded by different periods of a static pattern. In particular, the response in V1 and V2 was strong to a reversal when preceded by two static video frames. However, in V1 the response to a reversal was all but extinguished when preceded by static and reversal video frames while the response was still significant in area V2 for the same sequence. This rapid contrast gain control in V1 but not in V2 is a challenge to understand given the hierarchical nature of cortical organization. Either alternate pathways to V2 that bypass V1 carry the signal or the V1 to V2 pathway is of more than one type some of which are invisible to MEG and EEG recording methods.

STKE – rapid and intermediate contrast gain mechanisms: The STKE method provides the ability to create more complicated white noise experiments that allow for measuring kernel dynamics. To support the usefulness of the method we have demonstrated interesting aspects of contrast gain control of the visual evoked potential. A previous way to collect this type of evoked potential data is sometimes referred to as a double flash experiment. The STKE method is a very efficient method of collecting a large family of double flash experiments. The data presented was for a single patch but since this method is based on complete m-sequences it is readily expanded to multiple stimulus patches. By applying cortical constraints to isolate visual areas and applying the STKE method, it will be possible to evaluate the role of different visual areas in contrast processing. The method provides a complimentary approach to conditioned kernels to study rapid contrast gain while it also reveals much slower contrast gain processes.

5.3 Summary of innovations and findings

Past researchers have struggled to identify signals from early cortical areas and met with limited success. Moreover, they generally lacked a method of validating the solutions. The problem is that early visual areas are spatially close such that different linear combinations of the true source signals are not distinguishable based on scalp recordings. This problem is often referred to as the rotation problem in reference to using principal component analysis to separate signals. Our approach brings several innovations that effectively solve this otherwise intractable problem and provides a method of internal validation.

1) The use of high temporal rate dense multi-focal stimuli minimizes activation of late visual areas, with fewer sources to model the easer the problem becomes. Moreover, multi-focal methods limit the rotation problem by having some stimulus locations where the sources from different areas have near orthogonal magnitudes.

2) The common time function assumption across stimulus patches effectively provides a template for isolating the response for stimulus patches with low SNR.

3) By constraining dipole source localization to the flattened cortical surface we are able to estimate the surface topography for each source, the time function then is derived by a linear least squares regression step. With so many source estimates from the dense multi-focal stimulus the potential locations of individual sources in V1 and V2 are very limited. The cortically constrained method reveals that the VEP for area V1 and V2 begin at about same time but with opposite polarities. Previously the C1 potential at around 75 msec was thought to be V1. Our results indicate that it

likely involves some combination of V1 and V2 responses. The use of different stimuli, tasks and recording sites across laboratories impact the relative contribution of V1 and V2 sources which may explain prior contradictory findings.

4) Source identification results typically lack a means of validating the solutions. This cortically constrained method provides an internal validation of the derived V1 and V2 source time functions by comparing the solutions across hemispheres each of which has unique cortical folding patterns. Since the data from each hemisphere were processed independently, finding similar time functions across hemispheres validates the solutions and confirms the utility of the over all method.

5) To improve source identification we demonstrated a method of combining MEG and EEG data. In the past EEG and MEG appeared to give different time functions. Using multi-focal methods and SVD we have shown that both types of recording actually do reveal the same temporal responses when the rotation problem is accounted for. The matching time functions extend over four diverse nonlinear kernels.

6) In a series of appendices we (i) develop a multi-focal SSVEP with complex SVD analysis, (ii) discuss assumptions of common time function (patch independence) and common topography (condition independence), (iii) provide a new empirical method of improving the forward model, and (iv) discuss the power of concatenating EEG and MEG and concatenating multiple stimuli including steady state and m-sequence noise. .

7) Finally, we demonstrate a powerful m-sequence method to study contrast gain control which we call STKE. When STKE is used in a multi-focal paradigm it will enable us to identify the cortical areas involved in rapid and intermediate speed contrast gain control. Another approach to revealing rapid contrast gain control is to look at the nonlinear condition kernel responses. For the subject with cortically constrained method of isolating the V1 and V2 response, the condition kernels indicate different gain control processes in areas V1 and V2. Moreover, additional multi-focal m-sequence data from MEG/EEG recordings also point to differential contrast gain control processes in different visual areas. This by itself is a significant finding, particularly since V2 does not exhibit the rapid gain control evident in V1. The method also points to another tool for solving the rotation problem. Namely, if the finding of different gain control processes in V1and V2 is verified in more subjects, this could provide an alternative means to solve the rotation problems without resorting to the fMRI based cortical constrained technique. The promise of using VEP methods to study interactions between early visual areas may soon become a reality.

6. APPENDICIES

6.1 APPENDIX 1 - Multi-focal SSVEP and complex SVD

This Appendix extends the steady state VEP method to a multi-focal paradigm plus the application of complex SVD to factor the data, consistent with our m-sequence based methods. The steady state VEPs can be used to augment m-sequence methods for isolating early visual processing. It has the very strong advantage that the basic steady state response is captured by a just one pair of numbers for each harmonic. For checkerboard type stimuli where the negative and positive stimuli look alike, the odd harmonics vanish. Thus for a 5 Hz stimulus, only the 10 and 20 Hz harmonics would provide significant responses. Each harmonic is characterized by two numbers, amplitude and phase. The fact that only four numbers need to be estimated implies that good signal to noise values can be obtained in a relatively short recording time. This gives steady state stimuli a decided advantage when exploring nonlinear interactions among visual areas. The challenge is to do good source localization of the responses, so that idiosyncratic, reliable differences between areas can be identified. Tools for achieving that capability are discussed in this and the following Appendix.

While figure 1 includes 192 patches a more typical dartboard stimulus might contain four rings and eight spokes for a total of 32 patches, identical to the EEG/MEG stimuli discussed in Section 3. To increase the number of stimulus conditions each patch can, for example, be presented to the two eyes separately and each patch can, for example, contain two spatial frequencies. The need for 32 patches is to make use of the folding fingerprint of early visual areas to disambiguate multiple sources (V1/V2/V3). The need for multiple alternative conditions at each patch (two eyes/two spatial frequencies/or two contrasts) is needed to isolate independent sources by providing independent responses with the same topography. Two spatial frequencies (say 2 and 6 c/patch) will be important because of evidence that an important difference between striate and extrastriate cortex is that former has substantially higher response to high spatial frequencies than the latter. In Sections 3 and 4 we also discussed contrast processing approaches for distinguishing visual areas.

In the following analysis we consider runs of 500 sec in duration so that the frequency resolution is 0.002 Hz. The long runs are done by splicing, with overlap, five 100 sec segments. To frequency tag 32 patches, 2 eyes and 2 spatial frequencies in each patch we need 32*2*2 = 128 temporal frequencies all going simultaneously. Using counterphase oscillating stimuli the main response will be at the sum and difference frequencies. One problem is how to pack the tagging frequencies into a relatively small range of frequencies where the response is strong and similar to each other. Let us begin with a single patch where we need to tag four frequencies for the two eyes (R,L) and two spatial frequencies (1,2). As one of many possible examples (including frequencies whose second harmonic is outside the alpha range) let us take the four frequencies to be 5 (R1), 5.1(R2), 5.25(L1) and 5.45 (L2) Hz. The ten possible sum frequencies, combining all pairs of the four fundamental frequencies are: 10 (R1), 10.1(R12), 10.2 (R2), 10.25 (RL1), 10.35(R2L1), 10.45(R1L2), 10.5(L1), 10.55(RL2), 10.7(L12) and 10.9(L2) Hz. The frequencies for the 32 patches corresponding to the 5.0 Hz stimulus could be:

$$f(5.0, p) = 5.000 + .002*(p-1) \text{ Hz} \quad \text{for p=1 to 32}$$

The frequencies for the 32 patches corresponding to the 5.1 Hz stimulus would be:

$$f(5.1, p) = 5.100 + .002*(p-1) \text{ Hz}$$

The second harmonic response to 5.0 Hz would be at frequencies of $2*f(5.0, p)$ (e.g. 10.000, 10.004, etc). The cross modulation sum frequency of the 5.0 and 5.1 Hz stimuli would have frequencies $f(5.0) + f(5.1)$ Hz (e.g. 10.100, 10.102, 10.104, etc.). Thus just the 32 patches and the four conditions produce $10*32 = 320$ unique second harmonic (sum) frequencies. Care taken in assigning patch frequencies would allow one to pick up summation frequencies for neighboring patches (e.g. 10.002, 10.006, 10.010, etc.) that could double the number of measured frequencies. These interactions are useful for measuring center-surround and figure-ground interaction. Unwanted neighbor interactions could be avoided by introducing a very small gap between adjacent patches. In order to have unique *difference* frequencies as well as sum frequencies a quite different scheme is needed since the different patches would have the same difference frequencies. We are not interested in difference frequencies because they occur at low temporal frequencies where the noise is extremely large.

So far we have discussed the sum frequencies such as $2f_1$, $2f_2$ and f_1+f_2. The Regans[41,42] have demonstrated highly significant ultra-narrow higher harmonic responses. Of interest to us here are combinations such as $3f_1-f_2$ and $3f_2-f_1$. These would be seen at the unique frequencies of 9.9, 10.3, 9.75, 9.45, etc. Hz. There would be other combinations seen at frequencies that overlapped several sum frequencies. If the unique frequencies indicated substantial response, then a different scheme of assigning frequencies would be needed, possibly involving a reduction in the number of base frequencies. The statistical significance of each frequency component is evaluated using the Victor & Mast's tcirc statistic.[43,44]

Complex SVD

The SVD algorithm often plays a critical role in VEP analysis by dramatically reducing the number of free parameters, equivalent to a powerful reduction of noise. The SVD factorization in the frequency domain is:

$$V(ep, a) = \Sigma_c E(ep, c) F(c, a) \tag{A.1}$$

where e, p, a, and c are the electrodes ($N_e= 128$), patches ($N_p= 32$), alternatives ($N_a= 10$), and components ($N_c= 2$, first two account for most of the variance). Note that V and F are complex and E is real. The eigenvalue that represents strength of the component is grouped with F, so that E is . To keep E real one takes the imaginary part of the complex matrices and concatenates them into more columns. By number of alternatives N_a, we mean the number of types of stimuli at the same stimulus location. In the dichoptic case there were two eyes and two spatial frequencies, giving 10 individual responses for second harmonic sum frequencies (four $2f_i$, and 6 ($f_i + f_j$) see above for a more detailed counting of the alternatives in the dichoptic case). If they were done at five spatial and temporal frequency combinations the total number of alternatives would be $5*10=50$. That is, we assume the topographies ($E(ep,c)$) are the same across different experiments. For third and higher harmonics there would be an explosion of more terms.

The main assumption that we are making is that because all the patches differ by less than 0.05 Hz in our example so the phase shift in F(c,a) is the same across patches. We allow the phase to depend on the specific alternative and

component. The total numbers of data points in V are $N_eN_pN_a2$. The factor of two is because V is a complex number representing phase at each electrode for each patch. The total number of parameters on the right side of Eq. A1 is: $(N_eN_p + 2N_a)N_c$. Since $N_eN_p >> N_a$ the ratio is approximately $2N_a/N_c$. If we only consider the second harmonic terms $N_a = 10$ and if two components account for most of the variance then SVD is gives a ten-fold reduction in number of parameters. Equal in importance to this reduction in number of parameter is the conversion $N_a = 10$ complex valued topography maps for each patch to $N_c=2$ real valued maps. The challenge of dealing with complex valued maps is thus removed. I'm not quite clear in how to count the number of alternatives. Is it 10 or just 6, because four of the eye conditions are expected to be the same):

> 2nd harmonic of single eye for Freq 1; R1, L1
> 2nd harmonic of single eye for Freq 2; R2, L2
> Interaction of two frequencies: R12, L12
> Freq/eye interaction: R1L2, L1R2
> Eye interaction at Freq 1: RL1
> Eye interaction at Freq 2: RL2

That is, our conjecture is that in Eq. A1, F(c, R1) = F(c, L1) etc. Also one expects the freq interactions (R12, L12, R1L2, L1R2) to be small. These are nice checks.

Another way to look at this is that given the E(ep, c) topographies from <u>m-sequence</u> runs we just need 2(components)*10 complex numbers for F(c,a) to fit the SSVEP data (if the number of components doesn't change).

6.2 APPENDIX 2. A rotation problem associated with the common time function assumption and the common topography assumption.

It is useful to compare two very similar looking formulas. The first expresses the scalp voltages in terms of the underlying sources:

$$V(e, p, t, k) = \sum_{s=1}^{N} F_{true}(e, p, s)\, T_{true}(s, t, k)\ + noise(e, p, t, k) \qquad (A2.1)$$

where V, the raw data, is a function of electrodes (e), patches (p), time (t) and conditions (k). The second equation is the SVD output discussed in Section 3 in connection with our EEG/MEG data. For SVD, the data must be grouped into a two-dimensional matrix as follows:

$$V(ep, tk) = \sum_{c=1}^{N} E(ep, c)\, T(c, tk)\ + noise(ep, tk) \qquad (A2.2)$$

SVD applied to our multi-focal data shows that with just two components (N=2) the sum of squares of the signal (first term of Eq. A2.2) accounts for more than 80% of the sum of square EEG response for the 192 patch EEG data and for the 32 patch MEG data (see figure 7). The 32 patch EEG data in figure 7 was slightly noisier. Note that we have incorporated the eigenvalue S(c) of Eq. 3.1 into the topography (E) term. The low value of N is likely due to our use of relatively rapid, high spatial frequency, steady state stimulation. The finding that N=2 from SVD does a good job in accounting for the data, encourages us to hypothesize that N=2 is also appropriate for Eq. 1 that is based on the true source generators. Although the limitation to a small number of sources isn't relevant to the analysis of this Appendix, it is relevant to solving the inverse problem of estimating the time function, T_{true} and the forward model F_{true}. This Appendix explores the question of whether the small number SVD components places strict limits on the number of sources and limits on the separability of the sources. In particular we ask whether the low number of components provides evidence for a common time function assumption and a common topography assumption. The common time function assumes that the true time functions are independent of the patches involved. The common topography assumption is that the scalp topography from a given patch in a given visual area doesn't depend on the stimulus in that patch (assuming the stimulus fills the patch uniformly in all stimulus conditions). In order to examine the consequences of these assumptions we start from a more accurate equation where the constancy assumptions are not made:

$$V(e, p, k, t) = \sum_{s=1}^{3} F_{true}(e, p, k, s)\, T_{true}(s, p, k, t)\ + noise(e, p, k, t) \qquad (A2.3)$$

We take the summation to be over three sources. Although SVD indicated two sources it is good to be cautious and examine whether the 3rd source has signal. The new element in this form of the equation is that we have allowed the true time function to depend on the patch and the true topography to depend on the condition. For example, suppose that for s=2 (V2), the time function T_{true} depends on patch eccentricity. Similarly suppose that the V1 topography (s=1), F_{true} depends on stimulus strength or spatial frequency, represented by k. Alternatively, if a patch is surrounded by patches with stimuli moving upward it is possible that F_{true} will be shifted slightly (motion capture). If that patch happens to lie on the steeply curved part of a sulcus or gyrus, then the topography could change substantially since a small shift in cortical location can lead to a small rotation in the cortical surface normal leading to a large change in scalp topography.

In order to consider the consequences of small violations of the common time and common topography functions we take the novel step of applying SVD to expand both F_{ture} and T_{true}. :

$$F_{true}(e, p, k, s) = \Sigma_{s1} F_1(e, p, s, s1) R_1(s1, k) \tag{A2.4}$$

$$T_{true}(s, p, k, t) = \Sigma_{s2} T_2(s2, s, k, t) R_2(s2, p) \tag{A2.5}$$

Eq. A2.3 can now be rewritten to look like Eq. A2.1:

$$VEP(e, p, t, k) = \Sigma_{s'} F(e, p, s') T(s', t, k) + noise(e, p, t, k) \tag{A2.6}$$

where $\qquad\qquad$ $F(e, p, s') = F_1(e, p, s, s1) R_2(s2, p) \tag{A2.7}$

and $\qquad\qquad$ $T(s', t, k) = T_2(s2, s, k, t) R_1(s1, k) \tag{A2.8}$

where s' is a combination of s, s_1 and s_2. The summation over s' in Eq. A2.6 is actually a sum over the three indices s, s1 and s2.

It is useful to consider the simplest case where the summations over s1 and s2 in Eqs. A2.4 and A2.5 are restricted to s1=s2=1. In that case there can still be a violation of the constant time function and common topography assumptions, but the summation over components still only goes over s'=1 to 3 since s'=s. If the time function had a patch dependence that could be adequately summarized by two components in Eq. A2.5 (s2 ranging from 1 to 2) then the number of components being summed over in Eq. 2.6 would double, going from 3 to 6.

The preceding paragraph showed that it is possible for SVD to show that there are just two components and yet there could be patch dependence of the time functions as well as condition-dependent topographies (the restricted case of s1=s2=1). Our findings on two observers, based on <u>independent</u> analysis of small groupings of patches (only one observer showing only eccentricity dependence was shown in Fig. 5) is that patch dependence of time functions seems to be negligible. No differences were found for right vs left hemisphere, nor for upper and lower fields (after the video raster was taken into account). We were expecting to find an eccentricity dependence but no such dependence was reliably found. This claim has much higher confidence for V1 where the time functions were less noisy than for V2.

Condition dependence of the topographies is also doubtful. One of the most powerful illusion of perceived position shift is achieved in a motion induced shift paradigm. Whitney used fMRI to measure the cortical correlates of the shift and found it to be extremely small. In fact the very small shift that he found was in the opposite direction of the psychophysical shift, and attributable to a second order effect.

In summary, we believe that for purposes of using the algorithm to be discussed in Appendix 3, Eq. A2.6 gives an excellent representation of the true time functions and topographies even with a summation over a small number of components. However, it is always wise to carry out the comparisons with separate groups of fewer patches such as was done hemi-rings in Fig. 5 as a double check on whether the time function changed with eccentricity. .

This Appendix has identified a subtle rotation-like problem that can affect the true source expansion of Eq. A2.1. It is important to distinguish this subtle common time (or topography) rotation from the not at all subtle rotation problem that afflicts SVD. Eq. A2.2 can be rewritten in terms of arbitrarily rotated functions:

$$V(ep, tk) = \sum_{c=1}^{N} E_R(ep, c) T_R(c, tk) + noise(ep, tk) \tag{A2.9}$$

with
$$E_R(ep, c) = \sum_{c=1}^{N} E(ep, c') R^{-1}(c', c) \tag{A2.10}$$

and
$$T_R(c, tk) = \sum_{c'=1}^{N} R(c, c') T(c', tk) \tag{A2.11}$$

where R is an arbitrary rotation matrix that of E and T that leaves V unchanged. The rotation ambiguity provides a fundamental restriction on connecting the SVD derived topographies and time functions to the underlying sources. However, as Hood has argued and as we have discussed in Section 2, it might well be that at least for V1, in many individuals the first SVD component is close to the V1 component because the V1 source is the strongest so that it contributes the most variance, the SVD criterion for ordering components.

6.3 APPENDIX 3.   Using a time function projection to improve the estimate of the forward model.

This Appendix considers the question of how one could improve on the standard estimate of a forward model corresponding to a dipole in a visual area (say V2) responding to a small stimulus patch. To clarify the problem and its solution, it will be useful to start with a much simpler problem where one can avoid the complexity of multiple interacting dipoles. A good place to begin is with the cortical response to median nerve stimulation by electrical stimulation of the wrist. There is evidence that the very earliest cortical response is localized to a single small patch of somatosensory cortex, a single dipole. MRI and fMRI informs one of the location of brain activity. The EEG/MEG measurements give the scalp topography. There is no need to do a fancy boundary element calculation to get the forward model connecting the dipole location to the scalp topography. It is done empirically! We propose to do the same for early visual areas. The problem with the visual system is that over the past 50 years it has been controversial whether there is ever a moment when a single dipole is active. In fact some have argued that the first measured response is an extrastriate response (see Di Russo[24] for a review). Our data (Fig. 6) shows that the striate and extrastriate responses start at about the same time. Thus one can not simply look to the early response to obtain the empirical forward model. However, knowing the time functions from the early visual areas does give one the power to obtain the empirical forward model, as we now show using simulations.

Section 2 outlined our approach for a cortically constrained dipole fit to the EEG data. The multifocal stimulus we use is essential to our approach. The dartboard target with a large number of small patches, each flickering very rapidly, produces strong responses from early visual areas. However, later visual areas are not able to respond to the rapidity of the changes in the closely packed regions of space. This is consistent with our finding[19] that SVD analysis of the section 2 EEG data reveals two dominant components that account for 75% of the signal variance, similarly Fig 7 of section 3 shows 2 components accounting for 83% of the MEG response. A 3-shell forward model was used to estimate the V1 and V2 topographies for each patch. After a few iterations of shifting the dipoles (up to a total of 3.5 mm) for each patch and visual area on the flat map, the sum of square error was minimized and those dipole locations were used to generate an estimate of the topographies for each patch in each visual area (the forward model). Linear regression was then used to estimate the time function by multiplying the VEP data by the pseudoinverse of the forward mode (see Section 2 for details).

The new item that we would like to introduce is that now that one has an estimate of the time functions, one could do linear regression again, but this time use the time functions as the independent variable to refine the estimate of the forward model. In terms of equations the scalp potential or field is given by:

$$V(ep, tk) = \Sigma_s F_{true}(ep, s) T_{true}(s, tk) + noise(ep, tk) \tag{A3.1}$$

The algorithm of Section 2 gives the estimated time functions, presented in Figs. 5 and 6.

$$T_1(s, tk) = \Sigma_{ep} pinv(F_{fmri}(s, ep)) V(ep, tk) \tag{A3.2}$$

where $F_{3-shell}$ is the forward model based on dipole locations estimated from fMRI with orientations estimated from MRI. Note that we are adopting a convention of showing the ordering of pseudoinverse indices as they are after the pseudoinverse. The Matlab pseudoinverse function is used to carry out the linear regression $(pinv(A) = inv(A^T A)A^T)$.

We can improve the forward model so that it approaches the true forward model by iterating this process:

$$F_1(ep, s) = \Sigma_{tk} \, V(ep, tk) \, pinv(T_1(tk, s)) \tag{A3.3}$$

where F_1 is the improved forward model. Given an improved forward model why not get an improved estimate of the time functions?

$$T_2(s, tk) = \Sigma_{ep} \, pinv(F_1(s, ep)) \, V(ep, tk) \tag{A3.4}$$

And the improved time function lets us get a second order improvement on the forward model:

$$F_2(ep, s) = \Sigma_{tk} \, V(ep, tk) \, pinv(T_2(tk, s)) \tag{A3.5}$$

If this sounds totally crazy and unfamiliar that is because we have a very unusual situation of symmetry between the number of "independent" and "dependent" variables. As was discussed in Appendix 2, the data has the form, V(e, p, t, k) with about 100 electrodes (e), 48 patches (p), 100 time points (t) and about 50 conditions (k). Based on our evidence for a common time function using our multi-focal stimuli, the four dimensional data was concatenated in Eq. A3.1 to become the two-dimensional matrix, V(ep, tk). Although this looks like an approximately 5,000 by 5,000 matrix (see Appendix 4 for more details, including the integration with MEG) there are actually many fewer <u>independent</u> samples. It is well known that the electrodes and time functions are highly correlated. Furthermore, the cortical folding, although severe, still leaves substantial correlations between patch topographies. The various experimental conditions, such as the differences in gain control and temporal responses between different areas based on stimulus manipulations also are correlated. Thus we expect the number of truly independent samples would be reduced from about 5,000 rows and columns to roughly about 500 independent ep and tk variables.
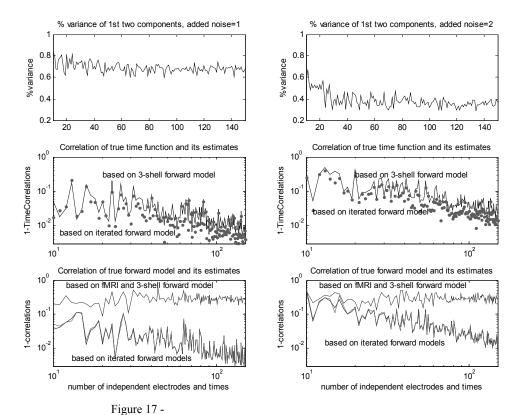
To determine whether our unusual back and forth linear regression makes sense we carried out simulations using the following Matlab program with the results shown in Fig. 17. The abscissa is N the number of independent electrode/patch values, that is taken to be the same as the number of time/conditions. We let N vary from 10 to 150, a range lower than the 500 independent samples of ep, and tk in our data. Our more than enough samples should allow us to subdivide our data to as part of our validation tests and patch independence tests as in the hemi-ring comparisons of Fig. 5. To guarantee independence in the simulations we take both the true topographies and time functions to be Gaussian random variables. Our choice of two sources and noise levels for the simulations are based on our findings across eight subjects[11] and the findings of ourselves and others[19,34] concerning how much of the variance is accounted for by the first few components.. In order to simulate the misspecification of the 3-shell estimate that starts the iteration in Eq. A3.2 we have degraded the true topography by adding noise. The Matlab code for the simulations is:

noise is set to 1 or 2 for the right and left panels of Fig. 17

```
for N=10:150                          % number of independent electrodes and time points
  F=randn(2,N);                       % the true forward model for two sources
  T=randn(2,N);                       % the two true temporal response
  V=F'*T+randn(N,N)*noise;            % electrode voltages plus noise
  [E,S,TT]= svd(V,'econ');            % do SVD
  Sdiag=diag(S)'.^2;                  % variance of each component
  variance(N)=sum(Sdiag(1:2)/sum(Sdiag));% sum of 1st 2 variance
  F3shell=F+randn(2,N);               % 3-shell estimate of F (note the noisy misspecification)
  T1=Ffmri'\V;                        % temporal response from algorithm of Section 2
  F1=V/T1;                            % improved estimate of forward model
  T2=F1\V;                            % improved estimate of time function
  F2=V/T2;                            % Do another iteration of forward model
  Cor=corrcoef([T(1,:)' T1(1,:)' T2(1,:)']); % correlations of time functions
  CorallT(N,:)=[Cor(1,2) Cor(1,3)]    % correls of true T to two estimates
  FF=[F(1,:)' F1(:,1) F2(:,1) Ffmri(1,:)'];
  Cor=corrcoef(FF);                   % correlation of true F to the three estimates
  Corall(N,:)=[Cor(1,2) Cor(1,3) Cor(1,4)]
end
```

The output of this program is plotted in Fig. 17. The abscissa in all the panels is N, the number of samples of ep and of tk. That is the VEP data is an N by N matrix. The left and right groups of panels are additive noise with a standard deviation of 1 and 2 respectively added to each element of VEP data. The top two panels show the percent of the variance accounted for the leading two SVD components. For additive noise of 1 and 2 the %variance accounted for is about 70% and 40%. As N increases somewhat less of the variance is accounted for because there are more noisy components. Our data tends to lie around this range of variance accounted for by 2 components.

The middle pair of panels (fig. 17) shows 1 minus the correlation of the estimated time function as compared to the true time function. By plotting 1 - correlation on log-log axes we are able to see what happens as the correlations get close to 100%. The line is for the time function estimated using the 3-shell topography. The dots are for the time function using the improved iterated forward model. For added noise = 1 and N=100 the correlations are about 98.6% and 99.1% for the 3-shell and the iterated time functions. The iterated time function has a few outliers that are as 'bad' as the 3-shell prediction. For noise=2, the correlations are around 95.7% and 97.4% respectively. The iteration thus enables an approximate halving of the correlation deviation from unity. The bottom panels show the correlation of the true topographies with the topographies estimated either by the fMRI/MRI/3-shell model, or by the first and second iterations from Eqs. A3.3 and A3.5. For the left panel (noise=1) the 3-shell correlation is 70% while both the 1st and second iterations have correlations with the true forward model of 99.1%! For the right panel the respective correlations are 70% and 97.4%! Note that the additive noise does not add extra error to the 3-shell estimates of topography since that estimate was independent of the VEP data.

Our simulations show that the iteration approach can give a modest improvement in estimates of the time functions and a dramatic new method for improving estimates of the forward model as long as one has some assurance that there are enough independent samples in ep and tk and assurance that only a small number of components are dominating the response. We are confident that both of these conditions are realizable with the type of multi-focal stimulation that we have been using.

Figure 17 -

6.4 APPENDIX 4. Generalized SVD for combining multifocal transient and steady state responses.

Our basic data are scalp voltage topographies at many electrodes (e) as a function of time (t), in response to a wide variety of stimulus conditions (k) at multiple visual stimulus patches (p) arranged in a dartboard fashion. As discussed in Appendices 2 and 3, the connection of the scalp voltages to multiple cortical sources is given by:

$$V(ep, tk) = \sum_{s=1}^{3} F_{true}(ep, s) \, T_{true}(s, tk) + noise(ep, tk) \qquad (A4.1)$$

Appendix 2 discussed the common time function assumption (T_{true} being independent of patch) and the common topography assumption (F_{true} being independent of stimulus condition). Eq. A4.1 is easily confused with a similar looking formula for the output of the SVD algorithm, also discussed in Appendix 2. We distinguish the two by using the summation index, s, for the source formula (A4.1) as opposed to the component index, c, for the nearly identical SVD expansion.

This Appendix examines the practical issues of what is needed to enable the Appendix 3 algorithmic improvements over our previous methods used in Section 2. Appendix 3 outlined a method for obtaining an empirically based forward model that reduces the fMRI localization errors and the errors associated with inaccurate head models due to unknown skull and brain conductivities and anisotropies. However, the Appendix 3 algorithm required extensive EEG time function data that had sufficient numbers of independent samples to reliably carry out the linear regression needed for estimating the topographies for multiple patches. In this Appendix we consider the numbers of independent time samples to be collected.

Our multi-focal methods enable the number of rows of V and F to be very large. We can easily get more than 100 EEG electrodes (we ignore MEG for now) and 100 patches. Thus the number of rows of the data matrix could be $N_e*N_p = 10,000$. If one wanted to process the right visual field independently from the left for validation, then there would be 5,000 samples, a number used in previous sections. However, these samples are not independent since volume conduction in the brain correlates electrode activity, and nearby patches tend to have similar dipole normals so nearby patches tend to be correlated. To a first approximation let us guess that in response to a hemifield of active patches there are 500 independent samples.

We would like to get a comparable number of independent samples for the time functions, the columns of the data matrix. Each independent m-sequence kernel has about 100 samples from t=40 to 240 msec with 2 msec sampling. Of these 100 samples there could roughly be 10-20 independent ones. From each m-sequence multi-focal experiment there are about 6 nonlinear kernels whose strength is equivalent or greater than the [03] and [0123] kernels shown in Fig. 7. Thus from each 20 min run (for a 16 bit m-sequence) we extract about 100 independent time samples. In order to get 500 independent columns for the data matrix V, we would need about 5 independent conditions. To be safe on the independence requirement we would do 10 conditions. We now examine what these conditions and stimuli might be.

In order for the iteration scheme of Appendix 3 to work well we need to have stimuli for which the separate visual areas (V1, V2, V3, V4, MT) have different time dependencies. We showed in Figs. 8 and 9 that the rapid contrast gain properties of the V1 and V2 sources have different properties. The temporal frequency of oscillations are also different as well as the after-ringing. We anticipate similar differences in later visual areas. Figs. 10 and 11 (for the first two SVD components) show similar differences. Stimulus manipulations of color, contrast[17], dichoptic presentation, spatial frequency, etc. can be expected to provide further differences between areas. Since all these stimuli would be presented using identical dartboard layouts, it is expected that the forward model topographies, F(ep, s), would be the same for all stimuli, and the time functions could be concatenated giving a very large number of independent samples. The large number of independent samples was needed for the algorithm of Appendix 3.

In addition to all the m-sequence based stimuli to be concatenated in F(ep, s), one can also include the steady state multi-focal responses that were discussed in Appendix 1. The two big advantages of steady state stimuli are: 1) The responses to a single stimulus can be summarized by about 10 numbers. For example for a 4 Hz stimulus one can get responses at 4, 8, 12, 16, 20 and 24 Hz (at much higher frequencies the signal to noise ratio (SNR) is not great. At each frequency one can extract the amplitude and phase (or real and imaginary components of the fft). For counterphase stimuli the odd harmonics (4, 12, 20) vanish so there would be only 6 independent numbers). With all the response energy concentrated in fewer numbers than with white noise stimuli, the SNR is greatly improved, so that many more

different conditions can be tested. 2) Within a given patch we typically combine several overlapping stimuli. In Appendix 1 we mentioned dichoptic stimuli and different spatial frequencies. Further details using overlapping stimuli were presented in our 2006 SPIE paper.[19] The overlapping stimuli allow us to measure interactions between eyes and between types of stimuli. In recent studies Jian Ding and we have found that these interactions have different topographies than the self responses, so we expect them to be rich sources for independent time functions. As was discussed in Appendix 1, for four stimuli (four frequencies, f_i) at each patch there would be 10 combinations of $f_i + f_j$. Consider the case $f_1=4$ and $f_2 = 4.1$. The sum frequencies (near 2nd harmonic) would be 8, 8.1, 8.2 Hz. Near the 4th harmonic, there would be 16, 16.2, 16.4 Hz. As pointed out by Regan & Regan[41, 42] there are also important frequency responses near the 2nd harmonic at $3f_i - f_j$ which would be at 3.9 and 4.3 Hz in our example. Our group with Jian Ding have recently carried out a number of experiments on flicker and jitter responses in which each patch has a stimulus given by:

$$P(x, t) = c_{ped} \cos(g^*x) + c_{flicker} \cos(g^*x) \cos(f_i^*t) + c_{jitter} \sin(g^*x) \sin(f_j^*t) \qquad (A4.2)$$

where the frequencies f_i and f_j depend on patch as well as on the flicker vs. jitter condition. We found that not only do we get responses at the sum frequencies, we also get significant responses at the $3f_i - f_j$ frequency. In addition the topographies were found to be different across the different conditions. It has been long known that the SSVEP displays a very rich behavior as a function of spatial and temporal frequencies[45,46] with regions where increasing stimulus contrast produces a decreasing response and spatiotemporal regions with rapid changes. The obvious explanation for these rich and exotic, but reliable effects is that there are multiple generators of the SSVEP that have different contrast and frequency dependencies. This is precisely what we need for independence of the time functions discussed above.

In SSVEP experiments we suggest using long runs with breaks. A typical run might be 500 sec, which would be presented as five 100 sec segments. In doing these runs careful synchronization of the segments is needed. One of the important experimental techniques that is needed for obtaining ultra-narrow frequency tuning of the fft is that the observer needs to be in a stationary state. In a typical run the observer is not stationary, but is adapting to the stimulus whether it is in a 10 sec, 100 sec or 500 sec. In order to minimize this type of adaptation, we will have a preadapt period and even during rest periods between runs the subject will still maintain fixation while being allow to blink and move around. The SNR of a single signal frequency component is about the same for doing the one long run of 500 sec vs averaging five 100 sec runs. The benefit of the longer run is that we can pack the different conditions much closer in frequency space so that the response amplitudes and phase shifts will be very similar across patches (see Appendix 1, for our frequency packing scheme.

It is important to note that the steady state experiments would be concatenated together with the m-sequence kernels in the same Appendix 2 type analysis. The SSVEP stimuli would use the same dartboard layout and presumably the topographies for each source (V1, V2 etc) would be the same. Thus the steady state stimuli are simply to be thought of as extra time points, adding more columns to the data matrix V(ep, tk) and to the time matrix T(s, tk).

It is useful to consider the impressive data reduction achieved by Eq. 4.1 or by the comparable SVD equation. For the V(ep, tk) data that has N rows and columns there are $N^2 = 25*10^4$ data for N=500. For three sources the number of parameters are 3N*2 where the factor of 2 comes from the parameters for F and for T. The ratio of data to parameters is N/6 = 83. This very large ratio underlies the robustness of the approach in Appendix 3 for using the empirical data using small stimulus patches to improve the fMRI/MRI based forward model with all its uncertainties. That improved forward model could then be used in totally independent experiments using different stimuli and combinations of patches for isolating particular visual areas. We are eager to apply the improved forward model to the 192 patch data shown in Fig. 5 to see if the correspondence between right and left hemispheres can become even better.

REFERENCES

[1] Zhang, Z., Jewett, D.L. and Goodwill, G., "Insidious errors in dipole parameters due to shell model misspecification using multiple time-points," Brain Topograpy 6, 283-298 (1994).

[2] Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D. and Halgren, E., "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity," Neuron 26(10), 55-67 (2000).

[3] Hämäläinen, M.S. and Ilmoniemi, R.J., "Interpreting magnetic fields of the brain: minimum norm estimates," Med Biol Eng Comput 32(1), 35-42 (1994).

[4] Phillips, C., Rugg, M.D. and Friston, K.J., "Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints," NeuroImage 16(3), 678-695 (2002).

[5] Baker, S., Baseler, H., Klein, S.A. and Carney, T., "Localizing sites of activation in primary visual cortex using visual-evoked potentials and functional magnetic resonance imaging," Journal of Clinical Neurophysiology 23(5), 404-415 (2006).

[6] Sharon, D., Hämäläinen, M.S., Tootell, R.B.H., Halgren, E. and Belliveau, J.W., "The advantage of combining MEG and EEG: Comparison to fMRI in focally stimulated visual cortex," NeuroImage 36(4), 1225-1235 (2007).

[7] Baillet, S., Riera, J.J., Marin, G., Mangin, J.F., Aubert, J. and Garnero, L., "Evaluation of inverse methods and head models for eeg source localization using a human skull phantom," Phys Med Biol 46(1), 77-96 (2001).

[8] Ferree, T.C., Clay M.T., Tucker D.M., "The spatial resolution of scalp EEG," Neurocomputing 38(40), 1209-1216 (2001).

[9] Lütkenhöner, B., "Dipole separability in a neuromagnetic source analysis," IEEE Trans Biomed Eng 45(5), 572-581 (1998).

[10] Dale, A.M., Fischl, B. and Sereno, M.I., "Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction," NeuroImage 9(2), 179-194 (1999).

[11] Dandekar, S., Ales, J., Carney, T. and Klein, S.A., "Methods for quantifying intra- and inter-subject variability of evoked potential data applied to the multifocal visual evoked potential," Journal of Neuroscience Methods 165, 270-286 (2007).

[12] Klein, S.A. and Carney, T., "The usefulness of the laplacian in principal component analysis and dipole source localization," Brain Topography 8(2), 91-108 (1995).

[13] Slotnick, S.D., Klein, S.A., Carney, T., Sutter, E. and Dastmalchi, S., "Using multi-stimulus VEP source localization to obtain a retinotopic map of human primary visual cortex," Clinical Neurophysiology 110(10), 1793-1800 (1999).

[14] Baseler, H.A., Sutter, E.E., Klein, S.A. and Carney, T., "The topography of visual evoked response properties across the visual field," Electroencephalography and Clinical Neurophysiology 90(1), 65-81 (1994).

[15] Hood, D.C, Ghadiali, Q., Zhang J.C., Graham, N.V. and Wolfson, S.S., "Contrast-response functions for multifocal visual evoked potentials: a test of a model relating V1 activity to multifocal visual evoked potentials activity," Journal of Vision 6(5), 580-593. (2006).

[16] Slotnick, S.D., Klein, S.A., Carney, T. and Sutter, E.E., "Electrophysiological estimate of human cortical magnification," Clinical Neurophysiology 112(7), 4113-4125 (2001).

[17] Baseler, H.A. and Sutter, E.E., "M and P components of the VEP and their visual field distribution," Vision Research 37(6), 675-690 (1997).

[18] Schira, M.M., Wade, A.R., Kontsevich, L.L. and Tyler, C.W., "Geometric and metric properties of visual areas V1 and V2 in humans," Journal of Vision 5(8), 897a (2005).

[19] Carney, T., Ales, J. and Klein S. A., "Advances in multifocal methods for imaging human brain activity," SPIE, 6057, 605716-1-12 (2006).

[20] Sutter, E.E., "Imaging visual function with the multifoical m-sequence technique," Vision Research, 41, 1241-1255 (2001).

[21] Engel, S.A., Glover, G.H. and Wandell, B.A., "Retinotopic organization in human visual cortex and the spatial precision of functional MRI," Cerebral Cortex 7(2), 181-92 (1997).

[22] Fischl, B., Sereno, M.I., Tootell, R.B. and Dale, A.M., "High-resolution intersubject averaging and a coordinate system for the cortical surface," Human Brain Mapping 8(4), 272-284 (1999).

[23] Teo, P.C., Sapiro, G. and Wandell, B.A., "Creating connected representations of cortical gray matter for functional MRI visualization," IEEE Transactions on Medical Imaging 16(2), 852-63 (1997).

[24] Di Russo, F. Martínez, A., Sereno, M. I., Pitzalis, S. and Hillyard S.A., "Cortical sources of the early components of the visual evoked potential," Human Brain Mapping 15(2), 95-111 (2002).

[25] Horton, J.C. and Hoyt, W,.F., "The representation of the visual field in human striate cortex. A revision of the classic Holmes map," Archives of Ophthalmology 109(6), 816-24 (1991).

[26] Mehta, A.D., Ulbert, I. and Schroeder, C.E., "Intermodal selective attention in monkeys I: distribution and timing of effects across visual areas." Cerebral Cortex 10(4), 343-358 (2000).

[27] Dayan, P. and Abbott, L.F., [Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems], MIT Press, Cambridge, (2001).

[28] Zhang, X. and Hood, D.C., "A principal component analysis of multifocal pattern reversal VEP," Journal of Vision 4(1), 32-43 (2004).

[29] Huang, M-X., Song, T., Hagler, D.J., Podgorny, I., Jousmaki, V., Cui, L., Gaa, K., Harrington, D.L., Dale, A.M., Lee, R.R., Elman, J. and Halgren, E., "A novel integrated MEG and EEG analysis method for dipolar sources," NeuroImage 37(3), 731-748 (2007).

[30] Fuchs, M., Wagner, M., Wischmann, H.A., Kohler, T., Theissen, A., Drenckhahn R. and Buchner, H., "Improving source reconstructions by combining bioelectric and biomagnetic data," Electroencephalography and Clinical. Neurophysiology. 107(2), 93-111 (1998).

[31] Klein, S.A., "Separating non-linearities and multiplicative noise in contrast discrimination," Vision Research 46(25), 4279-4293 (2006).

[32] Klein, S.A., "A local measure for modeling contrast discrimination: Response to Kakov, Tsodyks, and Sagi," Vision Research 47, 2912-2917 (2007).

[33] Hall, S.D., Hooliday, I.E., Hillebrand, A., Furlong, P.L., Singh, K.D. and Barnes, G.R., "Distinct contrast response functions in striate and extrastriate regions of visual cortex revealed with magnetoencephalography (MEG)," Clinical Neurophysiology 116(7), 1716-1722 (2005).

[34] Maddess, T., James, A.C., Bowman, E.A., "Contrast response of temporally sparse dichoptic multifocal visual evoked potentials," Visual Neuroscience 22(2), 153-162 (2005).

[35] Klein, S.A., [Nonlinear Vision: Determination of Neural Receptive Fields, Function, and Networks], CDC press, Boca Raton, 109-170, (1992).

[36] Ding, J. and Sperling, G., "A gain control theory of binocular combination," Proceedings of the National Academy of Sciences of the United States of America 103(4), 1141-6 (2006).

[37] Ding, J. and Sperling, G., [Computational Vision in Neural and Machine Systems], Cambridge University Press, Cambridge, (2007).

[38] Menz, M., Menz, M.K., Sutter, E.E., "Fast adaptive mechanisms in the retina revealed with the multifocal ERG," Investigative Ophthalmology and Visual Science 45(2), U404 (2004).

[39] Sutter, E., "The fast m-transform: A fast computation of cross-correlations with binary m-sequences," SIAM Journal on Computing 20, 686-694 (1991).

[40] Bobak, P., Bodis-Wollner, I., Marx, M.S., "Cortical Contrast Gain Control in Human Spatial Vision," Journal of Physiology-London 405, 421-437 (1988).

[41] Regan, D. and Regan, M.P., "A frequency domain technique for characterizing nonlinearities in biological systems," Journal of theoretical Biology 133, 293-317 (1988).

[42] Regan, D. and Regan, M.P., "Objective evidence for phase-independent spatial frequency analysis in the human visual pathway," Vision Research 28(1), 187-191 (1988).

[43] Mast, J. and Victor, J.D., "Fluctuations of steady-state VEPs: interaction of driven evoked potentials and the EEG," Electroencephalography and clinical neurophysiology 78, 389-401 (1991).

[44] Mast, J. and Victor, J.D., "A new statistic for steady-state evoked potentials," Electroencephalography and clinical neurophysiology 78, 378-388 (1991).

[45] Tyler, C.W., Apkarian, P. and Nakayama, K., "Multiple spatial-frequency tuning of electrical responses from human visual-cortex," Experimental Brain Research 33, 535-550 (1978).

[46] Tyler, C.W., Apkarian, P., Levi, D.M. and Nakayama, K., "Rapid assessment of visual function – electronic sweep technique for the pattern visual evoked-potential," Investigative Ophthalmology and Visual Science 18(7), 703-713 (1979).