Comparison of Pattern VEPs and Preferential-Looking Behavior in 3-Month-Old Infants

Samuel Sokol and Anne Moskowitz

Studies of visual acuity in human infants between 1 and 6 months of age using the visual-evoked potential (VEP) and forced-choice preferential looking (FPL) have shown that acuity is one to two octaves higher by VEP estimates than by FPL estimates. In an attempt to study these differences, the authors obtained both VEP and FPL data from 26 3-month-old infants. VEP data were obtained with gratings of 0.31, 0.62, 1.25 and 2.50 cycles/deg, which were counterphase alternated at 2 Hz. FPL data were obtained for stationary gratings using either the method of constant stimuli or a staircase procedure. Our study revealed three major findings: (1) recordable VEPs can be obtained for spatial patterns that are below threshold by behavioral measures; (2) the use of different scoring criteria that yields comparable VEP and FPL group mean acuities does not yield a significant correlation between VEP amplitude acuity and FPL acuity for individual infants, probably because of the inherent "noise" in each technique; and (3) when VEP latency rather than amplitude is used to estimate acuity, there is a significant correlation between electrophysiology and behavior. Invest Ophthalmol Vis Sci 26:359-365, 1985

The course of visual acuity development in infants and young children has been measured with three techniques: the visual-evoked potential (VEP), forced-choice preferential looking (FPL) and optokinetic nystagmus (OKN). VEP studies have shown that an infant's visual acuity at 4 weeks of age is equivalent to a Snellen value of 20/200 to 20/400 and improves to 20/20 by 6 months of age.1−5 FPL and OKN studies have shown that at 4 weeks of age visual acuity is, on the average, 20/400 to 20/800 and improves to approximately 20/100 to 20/150 by 6 months.5 Thus, at 6 months, "VEP acuity" is nearly three octaves better than "FPL or OKN acuity."

There are several possible explanations for these differences. First, VEP and FPL studies use different criteria to estimate acuity threshold. Estimates of VEP acuity are usually based on some variant of a "zero microvolt criterion," ie, threshold is defined as the spatial frequency at, or slightly above, the point at which the brain elicits a 0 μV signal. FPL studies use a more conservative criterion to estimate threshold, eg, interpolation to the 70 or 75% correct point on the psychometric frequency of seeing curve. Secondly, there are differences in the temporal nature of the stimuli that have been used to obtain VEP and FPL data. VEPs are elicited with phase alternating stimuli and may arise from mechanisms that are stimulated by local luminance changes, contrast reversal or movement. On the other hand, FPL studies usually are carried out with stationary gratings so that mechanisms underlying luminance modulation, contrast change, and movement detection probably do not contribute to the infant's response. A third possibility is that there may be fundamental differences in the neural substrates that underly the VEP and FPL responses. The information available at the level of area 17 (which is enhanced by the use of VEP response averaging) may not be of any value to the infant in terms of eliciting a discrete motor response. Alternatively, FPL techniques may not be sensitive enough to reveal that this information is available to the infant.

The goals of the present study were: (1) to replicate the previously reported VEP and FPL group differences using the same infants in the same lab; (2) to determine whether or not the choice of different criterion levels contributes to these differences; and (3) to determine for each infant whether there is a correlation between acuity estimated by VEP amplitude and by FPL percent correct, and between acuity estimated by VEP latency and by FPL percent correct.

**Materials and Methods**

**Subjects**

Both VEP and FPL data were obtained from 26 infants (11-13 weeks of age). Informed consent was
obtained from each parent at the beginning of the study. Typically, two or three sessions were necessary to complete the testing of each infant. A session was terminated if the infant became inattentive, fussy, or sleepy. All sessions for a given infant took place within a 1.5-week period.

VEPs

The stimuli consisted of square-wave gratings generated on a black-and-white TV monitor. Four spatial frequencies were used: 0.31, 0.63, 1.25, and 2.50 cycles/deg (96, 48, 24, and 12 min of arc). The gratings reversed at a rate of 3.75/sec (1.88 Hz) and were presented in a circular 11-deg field. The mean luminance of the screen was 1.9 log cd/m², and the contrast of the stripes was 0.84.

Gold cup electrodes were used to record the VEP. The active electrode was attached to the scalp approximately 1 cm above the inion on the midline; one ear served as reference and the other as ground. The signals were led through a preamplifier with a bandpass of 1 and 50 Hz. The amplified signals were averaged by a software program containing an artifact rejection routine that examined each sweep and rejected sweeps containing artifacts, such as large d.c. shifts produced by head or body movement. During recording, either a single sweep or the cumulated response could be monitored on an oscilloscope.

The infant sat on the parent's lap and viewed the TV screen binocularly from a distance of 50 cm. An observer standing behind the TV monitored the infant's fixation and operated a remote control switch to start and stop averaging. Averaging was initiated when the infant was quiet and the reflection of the stimulus field could be seen in the infant's pupils. If the infant looked away from the screen, averaging was stopped and then restarted when the infant resumed fixation.

During each recording session, the four spatial frequencies were presented in random order. A control condition was presented at the end of each session. This consisted of switching the monitor to the TV mode without an incoming video program, resulting in visual "noise," and recording brain activity to 32 or 64 sweeps.

Typically, measurable records were obtained with either 32 or 64 accumulations. When no response was discernable at 64 sweeps, recording was continued until 128 sweeps had accumulated. The averaged responses were plotted with an X-Y recorder and stored on a floppy diskette for later retrieval and analysis. The latency of the first major positive wave (P₁) and the amplitude of the N₁P₁ component were measured.

Preferred-Looking Technique

The apparatus consisted of a wooden partition with two circular screens 11 deg in diameter, one to the left and one to the right of a red fixation light. The centers of the screens were separated by 44 deg of visual angle. A small peephole was located below the fixation light. During testing, the infant sat on the parent's lap 50 cm from the partition, centered between the two stimulus targets, and viewed the screens binocularly. The observer sat behind the partition and watched the infant through the peephole. Between trials, the observer flashed the fixation light to center the infant's gaze. On each trial, the observer made a forced-choice judgment as to which side contained the grating pattern based on the infant's behavior, including head and eye movements and facial expressions. The experimenter operated the two slide projectors, recorded the observer's responses and provided the observer with feedback as to the correctness of his or her responses.

Twenty-one of the 26 infants were tested, using the method of constant stimuli. The stimuli were slides of stationary black-and-white square-wave gratings of the same four spatial frequencies used for VEP recording: 0.31, 0.63, 1.25, and 2.50 cycles/deg. The luminance and contrast of the grating slides matched that of the TV monitor; 1.9 log cd/m² and 0.84, respectively. On each trial, a grating was paired with a blank field of matched luminance. The side of presentation of the grating was pseudorandomized, the only constraint being that each stripe width appear an equal number of times on the right and left side. There were two separate 50-trial blocks, each made up of 25 trials of each of two spatial frequencies: block 1—0.31 and 1.25 cycles/deg and block 2—0.63 and 2.50 cycles/deg. Within a block, the order of the two spatial frequencies was randomized. Half of the infants were tested with block 1 first and half with block 2 first. After the session, the percentage of correct responses was tallied for each spatial frequency.

Five infants were tested using a staircase procedure based on an up-and-down transformed response (UDTR) rule. Slides were placed in the slide trays in order of increasing spatial frequency, with two slides of each spatial frequency. Infants were tested over a range of four octaves starting with 0.75 cycles/deg. The average luminance and contrast of the grating slides were matched to that of the TV monitor used for VEP recording. On each trial, a grating was
paired with a blank slide of matched luminance. On a given trial, if the observer's response was correct, the slide trays were advanced toward higher spatial frequencies; if the observer's response was incorrect, the trays were moved to the first of the two slides of the next lower spatial frequency. The sequence was terminated when 10 reversals (excluding the first reversal) were completed. Threshold was taken as the mean of the spatial frequencies at which the 10 reversals occurred.

**Results**

Figure 1 shows VEP waveforms and FPL observer's percent correct as a function of spatial frequency for two infants. As spatial frequency increases FPL percent correct declines, VEP amplitude decreases and P1 latency increases. The data from both infants show measurable VEPs at spatial frequencies below a 70% psychophysical threshold. For example, Figure 1a shows VEPs at 1.25 and 2.5 cycles/deg, while the observer's percent correct for each of these spatial frequencies was below threshold (56%). In Figure 1b there is a VEP waveform at 2.5 cycles/deg that is clearly different from noise, yet the observer's percent correct—50%—was below threshold.

**Amplitude Estimates of VEP Acuity versus FPL Acuity**

Figure 2 shows VEP amplitude (N, P1) and FPL percent correct as a function of spatial frequency for six infants. VEP acuity was estimated from these amplitude data by extrapolating a linear regression line to 0 μV. The straight lines were fit by the method of least squares from the peak of each infant's amplitude-spatial frequency function. For example, if the amplitude was maximal at 0.31 cycles/deg, the line was fit to four data points (see infants EC, NE); if the amplitude was largest at 0.63 cycles/deg, the line was fit to the data points for 0.63, 1.25, and 2.5 cycles/deg (see infants BS, AC, MR, MS). Using these constraints, VEP amplitude acuity was estimated for 17 infants.

FPL 70% and 55% thresholds were estimated from 12 of the 17 infants by applying probit analysis to the psychometric function that were generated by the method of constant stimuli. The threshold level for the other five infants who were tested with the two-down, one-up staircase rule was approximately 70%.

Group mean acuity scores were calculated and two comparisons were made: First, the comparison that has been made between labs in previous studies, 0 μV VEP versus 70% FPL; second 0 μV VEP versus 55% FPL criterion. The results obtained with the traditional criteria (0 μV VEP, 70% FPL) produced a mean VEP acuity of 20/100 and a mean FPL acuity of 20/390, a significant difference of two octaves (P < 0.05). While our VEP acuity estimate is in agreement with previously published studies, the FPL acuity obtained from our group of infants seemed low. In reviewing the literature, however, we found the acuity range for 3-month-old infants was 20/125 to 20/360, putting our FPL findings only 0.12 octave lower than the data of Gwiazda et al (20/360), whose equipment we have reproduced.

† We were unable to estimate acuity using the extrapolation method from the remaining eight infants because three had a slope of zero, ie, there was no change in amplitude as a function of spatial frequency, and five did not produce reliable data for 2.5 cycles/deg gratings.
The 55% FPL criterion produced a mean acuity value of 20/110, which was not significantly different (0.14 octave) from the 0 μV VEP acuity of 20/100. Thus, by using a less conservative FPL criterion, agreement can be achieved between the two techniques. It is important to note, however, that these acuity values are based on group data. When each infant's VEP and FPL acuity estimates were compared directly, there was no correlation between the two techniques. In other words, even when the criteria were shifted to produce equivalent group acuities, i.e., 55% FPL versus 0 μV VEP, there was still no correlation between an individual infant's VEP amplitude and FPL acuity.

**Latency Estimates of VEP Acuity versus FPL Acuity**

Previous VEP studies have shown that latency is less variable than amplitude. Further, we recently have obtained data showing that the time course of VEP latency change to patterns with small (10-20 min/arc) elements is similar to the time course of acuity development as measured by FPL. On this basis we derived VEP acuity estimates from P latency. Figure 3 shows P latency and FPL percent correct as a function of spatial frequency for the same six infants whose amplitude data are shown in Figure 2. Straight lines were fit by eye to the linear portion of the latency function and VEP acuity was estimated using an arbitrary criterion of 140 msec. The mean acuity values were 20/460 for the VEP and 20/390 for 70% FPL, a nonsignificant difference of 0.24 octave. Figure 4 shows a direct comparison of acuity estimates from P latency (140 msec criterion) and from FPL 70% correct for each infant. Not only did the group acuity estimates agree, but there was also a significant correlation between the two measures for individual infants (r = 0.623; P < 0.01).

To avoid the issue of what VEP criterion latency to use, we also compared 70% FPL acuity with the absolute latency obtained from each infant for the 2.5 cycle/deg grating. Figure 5 shows that absolute VEP latency and FPL acuity are also significantly correlated (r = 0.509; P < 0.025).

**Discussion**

Our study reveals three major findings. First, within individual infants, recordable VEPs can be obtained from spatial patterns, which by behavioral estimates are indiscriminable from an unpatterned stimulus (Fig. 1a, b). Secondly, the use of different scoring
criteria adjusted to eliminate the previously reported VEP and FPL group differences yields no correlation between VEP amplitude acuity and FPL acuity for individual infants. Third, when VEP latency rather than amplitude is used to estimate acuity, there is a significant correlation between electrophysiology and behavior.

Our finding that VEPs can be recorded with spatial patterns below threshold may be due in part to the enhancement of the signal from the occipital cortex.

---

Fig. 3. FPL percent correct (closed circles) and VEP latency (open circles) as a function of spatial frequency for six infants.

Fig. 4. Comparison between VEP acuity using a P1 latency criterion of 140 msec and FPL acuity using a 70% correct criterion. There was a significant correlation ($r = 0.623; P < 0.01$). Closed circles: method of constant stimuli; open triangles: staircase method.

Fig. 5. Comparison between absolute VEP latency for 2.5 cycles/deg gratings and FPL acuity using a 70% correct criterion. A straight line was fit by computer using the method of least squares.
by computer averaging. That is, the experimenter, not the infant, benefits from the use of computer averaging to increase the signal to noise ratio of VEPs elicited by high spatial frequency gratings. For example, we often saw a large amplitude \( P_1 \) component during each sweep of the oscilloscope when low spatial frequency gratings (0.31 and 0.62 cycles/deg) were shown to an infant but not when higher spatial frequencies were presented. Thus, high spatial frequency gratings may not appear any more salient to the infant in a VEP paradigm than in a two alternative forced-choice paradigm, yet recordable VEPs may be obtained. A more appropriate comparison of VEP and FPL techniques might be a trial-by-trial “psychophysical” analysis of VEP waveforms. For example, an observer would be presented with two waveforms, one in response to a patterned stimulus, the other in response to visual noise, and be instructed to decide which waveform was obtained when the infant was in fact looking at the pattern. From these data an observer’s “frequency of seeing” curve for VEPs could be constructed that would be equivalent to the observer’s percent correct FPL curves.

Our second finding, that group agreement can be achieved between the two techniques when a more conservative FPL criterion is used, suggests that the previously reported VEP and FPL acuity differences were due in part to the choice of qualitatively different scoring criteria. However, in our judgement there are no equivalent criteria that can equate the two techniques. This issue can be avoided by varying a common parameter within each technique and measuring its effect on the infant’s acuity. For example, we are currently measuring VEP and FPL acuity in the same infants at two luminance levels. Under these conditions the same criterion can be used at each luminance level making the choice of criteria across techniques irrelevant. The question now becomes: Within each technique, what relative change in acuity, if any, occurs when luminance is varied, and is the change similar for the two techniques?

There are a number of possible reasons for the lack of a significant correlation between VEP amplitude acuity and FPL acuity. First, VEP and FPL acuity may arise from entirely uncorrelated neural substrates. This is unlikely for two reasons: (1) the cortical areas that evoke the infant’s behavioral (FPL) response receive input from the same cortical locations that are responsible for generating the VEP, ie, FPL acuity is downstream from VEP acuity, and (2) we find a correlation between VEP latency and FPL acuity, so it is unlikely that VEP and FPL acuities are dissociated completely. A second possible reason is that both techniques produced “noisy” data, which could lead to a poor correlation. For example, when VEPs are recorded, a large source of variance is the “nonstationarity” of the VEP, ie, the amplitude and phase may vary with time.15 This would in turn affect the goodness of fit of a regression line fit to the data obtained for a series of different check sizes. In addition, each infant’s FPL frequency of seeing curve is subject to variance due to shifts in the infant’s behavioral state and criterion shifts by the observer. While we have no estimate of the standard error or confidence intervals for each infant’s VEP amplitude–check-size function, probit analysis provides an estimate of uncertainty for each infant’s FPL data. The range of standard errors for the 12 infants whose data were analyzed by probit techniques was 0.35 to 1.16 octaves, with a mean of 0.60 octave. If, for example, an infant had a threshold of 20 min with a standard error of 0.60 octave, the 68% confidence limits for that particular infant’s acuity would be from 20/260 to 20/600 (X = 20/400), a relatively large range. The fact that we find a significant correlation when VEP latency and FPL acuity are compared favors this interpretation. Because latency is less variable,13 the variance envelope of the correlation matrix is reduced and a significant correlation occurs. Strategies that could be used to reduce the VEP amplitude and FPL variability include the use of larger number of trials, a greater number of stimuli and fast VEP sweep techniques.10,15,16

The present data and previous reports from this17–19 and other labs20–23 have shown consistently that VEP latency is a good electrophysiologic correlate of behavior. For example, during the first 5 years of life, the time course of the development of \( P_1 \) latency for pattern stimuli parallels the time course of acuity development as measured by FPL and OPL techniques.14,24,25 VEP latency measures also have been shown to correlate with infant behavioral spectral sensitivity.22,23 In addition, latency is sensitive to retinal blur,17 is abnormal in amblyopia16 and reflects the amplitude of accommodation in adults and infants.19 One advantage that latency offers over amplitude is that the absolute value (msec) can be used across sessions since it is not subject to the same degree of “nonstationarity” that produces amplitude variability. Future studies should be directed toward the use of VEP latency as a direct estimate of acuity.

In summary, both VEP and FPL acuities have been estimated from the same 3-month-old infants. When the previously used VEP and FPL criteria are used to estimate acuity, VEP amplitude acuity is “higher” than FPL acuity by two octaves. When a more conservative criterion (55%) is applied to the probit transforms of the infant FPL data, the group mean acuities are equal (20/100). There is, however, no correlation between VEP and FPL acuity for
individual infants, which is most likely due to the “noise” inherent in each technique and not to a dissociation of the two acuity measures. When VEP acuity is estimated by latency, we find a significant correlation between electrophysiology and behavior.

Key words: infant visual acuity, VEP, preferential looking

Acknowledgments
The authors thank Drs. Davida Teller and Velma Dobson for their helpful comments and suggestions.

References