The Effect of Contrast and Size Scaling on Face Perception in Foveal and Extrafoveal Vision

Dean R. Melmoth, Heljä T. Kukkonen, Pia K. Mäkelä, and Jyrki M. Rovamo

PURPOSE. To determine whether face perception can be equalized across the visual field by scaling size and contrast simultaneously.

METHODS. Contrast sensitivities were measured for detection ($N = 1$) and identification ($N = 2-8$) of a target face as a function of size (0.4°-10°) across eccentricities ($E = 0°-10°$).

RESULTS. In all conditions contrast sensitivity first increased and then saturated, as a function of stimulus size. Maximum sensitivity ($S_{\text{max}}$) decreased, whereas critical size (where $S = S_{\text{max}}/\sqrt{2}$) increased with eccentricity and set size ($N$). At each set size, sensitivities from all eccentricities could be equated by double scaling—i.e., translation in horizontal (size) and vertical (contrast) dimensions on log-log axes. Similarly, at each eccentricity, data from all set sizes could be superimposed using double scaling. Furthermore, all data could be superimposed onto the foveal detection curve when double scaled according to the equation $F = 1 + E/E_2^1 + \log N_1N_2 + E(\log N)/K$, where $r$ is horizontal or vertical. This equation incorporates the eccentricity ($E_r$) and set size ($N_2$), where contrast and size double, as well as the interaction term ($K$).

CONCLUSIONS. Double scaling superimposes data. Not only is this possible across set sizes or eccentricities separately, but by combining their effects, a function is provided that collapses all data to a single curve, explaining all performance variation across eccentricity and set size. Our results support the proposition based on numeral recognition that failures of spatial scaling across eccentricities may simply reflect the need for scaling both size and contrast. (Invest Ophthalmol Vis Sci. 2000;41:2811–2819)

S

stimulus magnification compensates for peripheral performance deficits in many visual tasks. Measuring performance across stimulus sizes to observe whether spatial scaling (i.e., horizontal translation) superimposes peripheral and foveal data has shown that performance in many visual tasks is qualitatively similar across eccentricities.1–7

A common measure for eccentricity-dependent performance is $E_2$, the eccentricity at which stimulus size must double to maintain foveal performance level.9 More demanding extrafoveal tasks have smaller $E_2$ values, indicating that stimulus size must double at a smaller eccentricity. $E_2$ values show a 100-fold intertask variation.9

Extrafoveal performance in some tasks is consistently poorer, irrespective of stimulus magnification.10–15 Spatial scaling cannot normalize low-contrast alphanumeric character recognition or high-contrast reading15 across eccentricities,10–12 but can equalize high-contrast character recognition16 and reading of nonmeaningful word strings.6 Therefore, scaling of both size and contrast is needed to superimpose data curves at all contrasts.10–12,14

Face recognition involving four front-posed faces is not spatially scalable.14 However, shifting the data in two dimensions (i.e., scaling both size and contrast) allows superimposition of all data.14 This double scaling is possible, because curves at all eccentricities have the same shape and also maximum sensitivity and size for performance saturation change. However, little is known of how eccentricity interacts with task demands in face perception. We therefore investigated the effect of the number of faces on the scalability of face recognition and $E_2$ values required, by measuring foveal and extrafoveal contrast sensitivities as a function of image size.

METHODS

Apparatus

Images were generated using an AM PC-5200 computer (Abingdon, UK) with VGA graphics board and presented on a monitor (Flexscan F503-M; Eizo, Ishikawa, Japan) with a frame rate of 60 Hz and resolution of 640 × 480 pixels. Pixel size was 0.047 cm. Average photopic luminance was 50 cd/m². To increase luminance resolution, color channels were combined using a video summation device.17 Attenuating the signal from two color guns before combination increased the palette of luminance levels from 8 to 14 bits, whereas the number of simultaneously displayable levels remained at 8 bits. The use of a 2 × 2 dither increased both by 2 bits.18

Stimuli

Photographs of eight male faces with three poses (front-on and 45° both left and right) were digitized and standardized by removing nonfacial features (hair, ears). Faces were stretched to a standard 100 × 130 pixels (4.7 × 6.1 cm on the screen).
Individual faces did not appear distorted: A naturally narrow face does not appear unusual when slightly widened. To ensure that subjects were recognizing the individual in the two-to-eight faces conditions rather than simply detecting unique local features, three poses of each face were interchanged randomly (Fig. 1B). Image root mean square (RMS) contrast values were equalized, and subjective brightness was matched by shifting average image luminance. All faces were equally detectable, and the contrast thresholds of all face pairs were equally discriminable.

For each face, we generated a series of image triplets (different poses) with contrast decreasing in steps of 0.1 log_{10} units from supra- to subthreshold. Stimuli were presented at 0°, 2.5°, 5°, and 10° eccentricity. For extrafoveal viewing, the fixation point was placed at the nearest edge of the stimuli, because fixation in the center would have caused the nearest edge to get increasingly close to the fovea when stimuli were magnified, giving the periphery an unfair advantage and confounding results. For foveal face recognition, performance is equal whether fixation is at the edge or the center (Mäkelä et al., unpublished data, 1999). Pilot studies confirmed that this was true for our stimuli (Fig. 2), and foveal fixation was therefore placed in the center of the face. The point of this study was to determine whether foveal fixation can be achieved in the periphery, and therefore we selected the location that would most accurately mimic natural behavior, given the artificial confines of the laboratory. To achieve the desired range of retinal sizes (0.4°–10°), viewing distance varied between 57 and 716 cm, and/or stimulus dimensions were halved on the screen (i.e., stimuli were merely magnified or miniaturized versions of each other).

![Figure 1](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933219/) Stimuli used for the study (A). Original stimuli were edited to remove nonfacial features and then standardized for height and width. Subjective brightness of the faces was matched by shifting the average luminance of the images while contrast was adjusted to be equal in terms of RMS contrast. Pilot studies showed that after this manipulation, all faces were equally detectable and all face pairs were equally discriminable. For all tasks, three poses of each face were randomly presented (B).

![Figure 2](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933219/) Pilot data confirming that foveal face recognition is equal whether fixation is at the edge or the center. To mimic natural face recognition as far as possible, given the artificial environment of a laboratory, the center of the face was selected. For extrafoveal viewing, however, eccentricity was measured from the closest edge of the stimuli, because measuring eccentricity from the center of the image would cause the nearest edge to move ever closer to the fovea, confounding results by giving the periphery an advantage.

**Procedure**

Subjects practiced until highly familiar with the procedures and stimuli. When no further improvement in performance was observed, the experiments began. Subjects fixated a black spot on the screen, or a green LED at greater eccentricities, and pressed a key to initiate a trial. Viewing was binocular. Trials with N = 1 comprised two 500-msec exposures (target and blank field) providing a detection task. There was only one exposure for other set sizes (N = 2–8) with subjects having to identify the individual presented.

Trials started at suprathreshold contrast, and if subjects responded correctly, contrast was reduced by a factor of 1.26. The second incorrect response initiated a staircase procedure with a four-correct-down/one-incorrect-up algorithm. Contrast thresholds obtained represented the probability level of 84%, 84%, 71%, and 59% correct for N = 1, 2, 4, and 8, respectively. Corresponding chance performance levels were 50%, 50%, 25%, and 12.5%. Threshold was calculated as the arithmetic mean of eight contrast reversals. RMS contrast sensitivity is the inverse of RMS contrast at threshold, defined as e_{RMS} = \sqrt{\epsilon/A}, where \epsilon is contrast energy at threshold and A is stimulus area in degrees squared. Stimulus energy is given as \epsilon = \sum c(x, y)p^2, where c(x, y) = |L(x, y) - L_0|/L_0 is local contrast at each image pixel, and p is pixel side length in degrees. L(x, y) refers to local luminance, and L_0 indicates average luminance.

**Goodness of Fit**

Goodness of fit can be described as

\[
G = 100(1 - e_{RMS})
\]
and,

$$e_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\log Y_{\text{est}} - \log Y_{i})^2} \quad (2)$$

where \(n\) is the number of data points, \(Y_{i}\) are the observed data, and \(Y_{\text{est}}\) are the estimates from the equation of least squares. This measure is appropriate for data on logarithmic coordinates (Figs. 3, 5, 7, 9). For data on linear coordinates (Figs. 4, 6, 8), relative differences between observed and estimated values were used.

$$e_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_{\text{est}} - Y_{i}}{Y_{i}}\right)^2} \quad (3)$$

### Subjects
Subjects DM (aged 24 years) and VJ (aged 22 years) had normal, or corrected to normal, vision. The Helsinki Declaration principles were observed.

### RESULTS
Figure 3 shows contrast sensitivities for detection of a face (\(N = 1\)) and identification from a set of faces (\(N = 2–8\)) as a function of stimulus height at 0°, 2.5°, 5°, and 10° along the horizontal meridian in the right visual field. For all set sizes and eccentricities, sensitivity first increased with stimulus size (height) and then saturated. The increase of recognition sensitivity with image height has been described as

$$S = S_{\text{max}} \left[1 + (b_{c}/b)^{2}\right]^{-m} \quad (4)$$

where \(S\) is sensitivity, \(S_{\text{max}}\) is maximum sensitivity, \(b\) is image height, \(b_{c}\) is critical height marking the saturation of contrast sensitivity increase. The value of \(4m\) indicates the slope of increase in double logarithmic coordinates, which is approximately 2 for both subjects in all conditions. Therefore, equation 4 in logarithmic form with \(m = 0.5\) was fitted separately to each data set. According to equation 4 with \(m = 0.5\), \(S = S_{\text{max}} \sqrt{2} \) at \(b = b_{c}\). Percentage values indicate goodness of fit.

Comparing curves within each frame of Figure 3 shows that critical size increased with eccentricity. In spatial scaling, peripheral data curves are traditionally shifted horizontally to superimpose onto foveal data. If the slope of increase and maximum sensitivities are the same, superimposition is successful, because the only difference is critical size.

Figure 3 shows that this is not the case, because maximum sensitivity decreases with increasing eccentricity. Thus, identification (\(N = 2–8\)) and even detection of a single face (with three poses) cannot be normalized across the visual field by spatial scaling alone. However, dependence of sensitivity on size is essentially identical in qualitative terms across eccentricities, because all curves have the same slope of increase and saturation rate. Thus, peripheral curves for each set size can be superimposed onto the corresponding foveal curve by scaling data in two dimensions: size and contrast. A deficit in contrast after size scaling has been reported by Strasburger et al.10–11 who found a hyperbolic relationship between size and logarithmic threshold contrast. Therefore, their data were not scaled with the traditional log–log method where sizes are divided by scaling factors. Therefore, no previous attempt has been made to simultaneously scale contrast and size in the same manner as size has been scaled in numerous earlier reports, nor are there any previous estimates of \(E_{2}\) values for contrast.

To scale the data, we took maximum sensitivities and critical heights and divided peripherally increasing critical

![Figure 3](http://iovs.arvojournals.org/pdfaccess.ashx?url=/data/journals/iovs/933219/ on 11/29/2017)
heights of each task by the corresponding foveal value, while foveal maximum sensitivity of each task was divided by the corresponding peripherally decreasing values. This provided the scaling factors that superimposed peripheral and foveal data. This is possible because all data curves, irrespective of eccentricity, have the same shape, and superimposing is therefore simply a matter of translation in two dimensions.

Figure 4 presents the scaling factors obtained. Irrespective of the number of faces\(^1\)–\(^8\) or type of scaling (horizontal or vertical), factors increased linearly with eccentricity. The gradients of the lines indicate how quickly scaling factors must increase to maintain constant performance with increasing eccentricity. The slope’s inverse, called \(E_2\), represents the eccentricity at which stimulus size\(^2\) (horizontal scaling) or contrast (vertical scaling) must double to maintain performance at the foveal level.

Scaling factors \((F_h, F_v)\) necessary to maintain foveal performance at any eccentricity \((E)\) are

\[
F_i = 1 + \frac{E}{E_{2i}}
\]

where \(i\) is either (h)orizontal or (v)ertical scaling. Scaling factor at \(E = 0\) is always 1, because foveal data are superimposed onto themselves.

We fitted equation 5 to each data set in Figure 4 and calculated the scaling factors for each eccentricity using the \(E_2\) values. On average, goodness of fit was 90% indicating that the lines describe the data of Figure 4 well. \(E_2\) values for spatial scaling \((E_{2h})\) and contrast scaling \((E_{2v})\) are shown separately for each task and subject.

Figure 5 shows the original data from Figure 3 double scaled according to equation 5. The curve fitted to the foveal data using equation 4 for each set size is included in each frame. All data superimpose and are described very well by the foveal curve.

Double \(E_2\) scaling thus compensates for extrafoveal increase in critical size and decrease in maximum sensitivity by comparing performance at each set size across eccentricities and scaling them to foveal performance (i.e., scaling data within each frame of Fig. 3). However, comparing curves of Figure 3 across set sizes at each eccentricity shows that critical size increased and maximum sensitivity decreased with increasing number of faces. Given the uniform shape of all functions, it was thus possible to scale data across set sizes at each eccentricity separately rather than scaling data across eccentricities for each set size separately, as in Figure 5.

Figure 6 shows scaling factors that superimpose set sizes \(N = 2\) to \(8\) onto the detection curve \((N = 1)\) at each eccentricity. Irrespective of eccentricity or direction of scaling (horizontal or vertical) scaling factors increased as a linear function of the logarithmic number of faces. Analogous to \(E_2\), we introduced an \(N_i\) value relating to the rate of scaling necessary to superimpose data of increasing set size. We thus obtained vertical and horizontal scaling factors for any number of faces using equation

\[
F_i = 1 + \log\frac{N}{\log N_{2i}}
\]

where \(i\) is (h)orizontal or (v)ertical scaling, \(N\) is set size, and \(N_{2i}\) is set size at which size or contrast must double to maintain detection performance level. Thus, instead of scaling extrafoveal data to the foveal \((E = 0)\) curve, equation 6 scales larger \((N = 2–8)\) set sizes (identification tasks) to the detection curve \((N = 1)\). As \(N = 1\) is superimposed onto itself, its scaling factor must be equal to unity, explaining the logarithms of equation 6.

Figure 7 shows the original data from Figure 3 replotted so that each frame represents a different eccentricity, rather than a different set size (Fig. 5), and scaled according to the \(N_{2i}\) and \(N_{2\text{th}}\) values, by using equation 6. The smooth curve is the detection function \((N = 1)\) for each eccentricity from Figure 3. Again, all data superimpose well, showing that simple quantitative shifts can compensate for all changes with set size.
All eccentricities were successfully scaled to foveal data (Figs. 4, 5) and all set sizes to N = 1 (Figs. 6, 7). This was possible, because all curves had the same shape, suggesting that all data could be scaled to a single function combining both eccentricity and set size. This function would superimpose all data onto the foveal detection curve (E = 0, N = 1). To reveal the function, critical size and maximum sensitivity values were divided as before, but instead of using the foveal function of each set size separately (E scaling) or the detection task of each eccentricity separately (N scaling), all functions were scaled to the foveal detection curve.

Figures 8A through 8D show the resultant scaling factors. Each scaling surface shows a series of scaling factors as a function of eccentricity resembling those in Figure 4 but separated in depth by the number of faces on logarithmic axis. Because all data are scaled to foveal detection, only this value is set at unity. Figures 8A through 8D show that double scaling is required after an increase in either eccentricity or set size, reflecting the decrease in maximum sensitivity and increase in

**FIGURE 5.** Original data after double $E_2$ scaling. Increasing extrafoveal size values are divided (for spatial scaling) by a factor based on the $E_{2h}$ values of Figures 4E through 4H, and decreasing RMS contrast sensitivity values are multiplied according to $E_{2v}$ values from Figures 4A through 4D to superimpose peripheral data onto the foveal curve at each set size. The smooth curves are the original functions fitted to the foveal data of each set size. After double scaling, all data follow this function very closely, as indicated by the goodness of fit values based on logarithmic RMS error. (A through D) subject DM; (E through H) subject VJ.

**FIGURE 6.** Scaling factors necessary to scale data from all set sizes ($N = 2–8$) onto detection data ($N = 1$) at each eccentricity separately. A linear function (equation 6) was fitted to each set of data. The inverse of the slope for each line of least squares represents an eccentricity-specific $N_2$ value, where vertical (A through D) and horizontal (E through H) $N_2$ values represent the set sizes at which stimulus parameters (size and contrast) must double to maintain the detection performance level ($N = 1$). After double scaling all data follow this function very closely, as indicated by the goodness of fit values based on logarithmic RMS error.
critical size seen in Figure 3. A scaling surface equation has more validity than the tailor-made values of Figures 8A through 8D or specific $E_{2}$ and $N_{2}$ values for each task and eccentricity, if that surface can explain all variation across eccentricities and set sizes. We suggested the following equation for all scaling factors ($F_i$):

$$F_i = 1 + E/E_{2i} + \log N/\log N_{2i} + E(\log N)/K_i$$  \hspace{1cm} (7)

where $E$ and $N$ are eccentricity and set size and $E_{2}$ and $N_{2}$ are the eccentricity and number of faces at which stimulus parameter, size (horizontal), or contrast (vertical) must double to maintain foveal detection performance. $K$ is a constant weighting the multiplicative interaction between set size and eccentricity: when either variable ($E$ or $N$) is small, the increase of scaling factor from foveal detection is small. When $E$ and $N$ both become large, however, they interact multiplicatively causing a much greater increase in scaling factor.

Figures 8E through 8H show the values predicted by equation 7 fitted in logarithmic form to the data of Figures 8A through 8D. Percentage values indicate the goodness of fit. Equation 7 represents a combination of equations 5 and 6 plus a term accounting for the interaction between eccentricity and set size. When $E = 0$ (foveal tasks) or $N = 1$ and thus $\log N = 0$ (detection tasks), equation 7 reduces to equation 6 or 5, respectively. Thus, $E_{2}$ and $N_{2}$ values of Figures 8E through 8H resemble the detection $E_{2}$ values (Figs. 4A, 4B) and foveal $N_{2}$ values (Figs. 6A, 6B), respectively.

Figure 9 shows the original data from Figure 3 scaled according to the surfaces of Figures 8E through 8H. Goodness of fit indicates the success of scaling. All data collapse to a single function, described well by the foveal detection curve, showing that equation 7 can account for all changes in performance across eccentricities and set sizes.

**DISCUSSION**

Our experiments on face perception equated sensitivities across the visual field by double $E_{2}$ scaling (i.e., size and contrast). This result obtained with three poses agrees with a report using four faces with one pose 13 but also extends the finding to include set size ($N = 1$–8).

Detection of geometric distortions of a high-contrast face is spatially scalable 21 whereas in the present study spatial scaling alone was not sufficient for any task, including detection. This is presumably due to low contrast and/or random presentation of three poses preventing the visual system using a single template.

The need for spatial scaling with increasing eccentricity indicates that the neural templates used for face perception suffer from positional uncertainty inherent in peripheral vision. 22–25 Stimulus magnification also takes care of reduced sampling, poorer resolution and larger receptive fields. The need for contrast scaling suggests an additional deficit of extraneous viewing. A study using four faces with a single front-on pose showed that decreasing $S_{max}$ with increasing eccentricity is due to reduced efficiency (Mäkelä et al., unpublished data, 1999), which can be compensated for by increasing contrast.

Although contrast sensitivity decreased with increasing set size at each eccentricity, all sensitivity curves had the same shape. Thus, an approach analogous to double $E_{2}$ scaling equated sensitivities across set sizes at each eccentricity ($N_{2}$ scaling), superimposing data for $N = 2$ to 8 onto the $N = 1$ curve. Face perception became more demanding with increasing set size, presumably because a greater number of face stimuli (and therefore nonorthogonal neural templates) results in the increased probability of a wrong choice. The success of double scaling implies that with an increasing number of faces the templates have two types of noise: positional inaccuracy,
which can be compensated by increasing image size and reduced efficiency, possibly through decreased signal-to-noise ratio of the template,\textsuperscript{26} which can be compensated by increasing contrast.

Combining $E_2$ and $N_2$ scaling showed that when eccentricity, set size, and their interaction are taken into account, all data can be scaled to a single function. The close match of the scaling surfaces to the observed factors (Fig. 8) and the goodness of fit of the collapsed data (Fig. 9) to the foveal detection curve validate the accuracy of $E_2N_2$ scaling of size and contrast.

Peripheral vision is affected by the crowding effect, where adjacent contours interact making stimuli very difficult to interpret.\textsuperscript{6,27 to 32} Thus, it is surprising that for face recognition, scaling contrast in addition to size is sufficient to compensate for all extrafoveal deficits. With word recognition, letter spacing has been increased in addition to magnification to reduce the crowding effect.\textsuperscript{6,30} Scaling size produces a downward shift in stimulus spatial frequencies, eventually disrupting recognition. Furthermore, although increasing letter separation reduces the magnification requirement, the overall size of words is larger, due to greater letter spacing.\textsuperscript{6} If scaling contrast reduces the need to increase size and spacing, the crowding problem, and that of trying to produce large images on a limited display, may be overcome.
Double scaling makes task comparison harder: if one task has a lower horizontal $E_2$ value, whereas another has a lower vertical $E_2$ value, which is affected more in extrafoveal vision? First, a distinction must be made regarding the interpretation of $E_2$ values. Spatial scaling reflects compensation for cortical factors, relating to the spatial grain of the visual field, whereas contrast scaling compensates for reduced efficiency and may therefore be task specific. It could be speculated that one of a few general horizontal $E_2$ values reflecting the known processing streams could be applied to all tasks, with task-specific contrast scaling added as required. This may help to explain the apparent 100-fold interset difference between spatial $E_2$ values found without contrast scaling. If double scaling is actually required, the sole horizontal shift that superimposes the lower, high-contrast, portions of the curves will be greater because in double scaling, data are superimposed through the shortest Euclidean distance.

This scenario is complicated by the significant effect (up to half the total scaling requirement) of the interaction term. It may be assumed that interaction increases with task difficulty, and thus interaction ($K$) could serve as a relative measure of difficulty for double-scaled tasks as $E_2$ does for spatially scaled tasks. Compensation for parameter interaction could, in fact, be compensating for the effect of crowding, which is known to increase with both eccentricity and, for example, the addition of flanking stimuli. In our study interaction increased with both eccentricity and set size. Size and contrast scaling firstly compensated for task difficulty but also eliminated the need for further manipulation to overcome crowding. It is therefore likely that the interaction observed and quantified here reflects extrafoveal crowding.

Success or failure of spatial scaling has been used as evidence of quantitative versus qualitative changes in processing across the visual field. The premise that failure of spatial scaling indicates a qualitative difference between foveal and extrafoveal processing was valid when stimulus size was the only dimension considered. However, the current results and those of Strasburger et al. proposed a "window of visual intelligence" in which fine details and low contrasts are processed: The more demanding the task, the narrower the window. Our results do not dispute this interpretation. Because of the huge demand on cortical resources, a logical evolutionary strategy would be to have a narrow specialized field, with the rest of the retina handling simple processing. The qualitative similarity between all data shows that sufficient double scaling would allow the periphery to process demanding tasks. Thus, our interpretation of the window is that, although this dichotomy exists in everyday life, it is quantitative not qualitative (at least for face recognition) and, given artificial manipulation to compensate for reduced resolution and contrast efficiency, performance can be equated.

Peripheral visual performance is far more complicated than $M$ scaling predicts. Strasburger and Rentschler observe that "spatial resolution is but one factor of the visual sense of form," whereas Chung et al. conclude that "size is not the only factor limiting reading speed in peripheral vision." The present study leads to a similar conclusion. Changes in resolution, crowding, and efficiency of contrast usage all play a role in extrafoveal performance. Scaling in multiple dimensions does increase the complexity of any extrafoveal model, but shows how alternative manipulations can enhance performance. This could benefit both theoretical study and practical performance-oriented application. The proposed shift from one-dimensional spatial scaling to double scaling mirrors the transition from $M$ scaling to spatial scaling, when it became apparent that the restrictions of predetermined $M$ scaling produced false-negative results for tasks that were actually spatially scalable. Again, the problem of scaling failures must be addressed by adopting a more versatile approach when comparing foveal and peripheral visual performance.

References

15. Hess RF, Field D. Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disarray? Vision Res. 1993;33:2665–2670.

Extrafoveal Face Perception and Set Size 2819