Published in final edited form as:
Eye Contact Lens. 2012 July ; 38(4): 252–259. doi:10.1097/ICL.0b013e31825aa879.

QUANTIFICATION OF GHOSTING PRODUCED WITH PRESBYOPIC CONTACT LENS CORRECTION

Pete S. Kollbaum, OD, PhD, **BoKaye M. Dietmeier, BS**, **Meredith E. Jansen, OD, MS**, and **Martin E. Rickert, PhD**
Indiana University School of Optometry, Bloomington, IN, USA

Abstract

Objectives—The defocused portion of the image obtained in wearers of bifocal and multifocal contact lenses often appears as a "ghost." Relatively few methods exist to quantify the ghosting perceived with lenses. The purpose of the current study is to validate and implement a questionnaire to help patients quantify the ghost images perceived with bifocal or multifocal corrections.

Methods—Ten subjects viewed simulated bifocal vision images displayed on a monitor. Images contained a focused and a defocused (ghost) component of a specific dimension (direction, position-offset, intensity, and focus). Using a test card, subjects identified the ghosting dimension level displayed on the monitor. An additional 54 presbyopic subjects wearing a multifocal correction monocularly viewed a well-focused stimulus and then compared the perceived image to that of the other well-corrected eye looking at the ghosting test card to quantify their visual experience of the 4 proposed ghosting dimensions.

Results—Regardless of ghost letter size and orientation, subject responses were within 1 rating unit of expected on >95% of all trials for all four dimensions when asked to directly match a single dimension of ghosting. With bifocal images containing random amounts of these four dimensions most response errors were also within +/−1. In presbyopes wearing a multifocal lens, the focus dimension was most strongly associated with overall ratings of ghosting.

Conclusions—Subjects can accurately and reliably report on ghost intensity, focus, direction and position-offset, and well-focused ghosts are most correlated with the overall perceptual saliency of ghosting.

Keywords

Bifocal Contact Lenses; Ghosting; Monocular Diplopia; Aberrations

INTRODUCTION

Visual acuity (VA) is routinely used as a surrogate measure of optical quality. For example, subjective refractions strive to maximize VA and refractive surgeries are often judged by the levels of uncorrected VA they deliver.[1–2] Additionally, VA is also often used as a surrogate measure of retinal image quality. For example, VA loss is often utilized to indirectly monitor the progression of macular disease.[3–4] However, in eyes utilizing bifocal or multifocal contact lenses, there is clear evidence that visual acuity is a poor indicator of

Pete Kollbaum, OD, PhD, FAAO, Indiana University School of Optometry, 800 East Atwater Avenue, Bloomington, IN 47405, phone: 812.856.0108, fax: 812.855.7045, kollbaum@indiana.edu.

Conflicts of Interest
The authors have no conflicts of interest to disclose related to this work.

visual quality in that patients with high levels of VA are often dissatisfied with their bifocal corrections.[5–8] Also, high contrast visual acuity has been shown to be limited in its ability to indicate the levels of higher order optical aberrations in an eye,[9] and may also have limitations in describing true quality of vision in many circumstances.[10]

The inability of acuity measures to adequately describe patient visual quality has motivated the development of methods for quantifying subjective visual quality directly[11] such as numeric rating scales (e.g. 0–100),[8] visual analog scales,[12] or some form of categorical (or Likert) scale.[13] These different types of scales have been reported to have similar capabilities for describing visual quality in the presence of small levels of spherical defocus.[14–15]

Although a visual quality scale may be effective at capturing the description of the impact of a one dimensional change in optical quality, (e.g. reduced spatial bandwidth due to increasing defocus),[15] it is clear that retinal images can simultaneously vary along multiple dimensions (e.g., spatial bandwidth remains high, but with reduced contrast) in the presence of defocus and higher order aberrations[16–20] and in the presence of bifocal optics.[21] Images produced by bifocal,[5, 7] multifocal,[8, 22] or intraocular lenses[23–24] can include both a focused image and a simultaneously present defocused or “ghost” image. Therefore, patients with such lenses may experience poor subjective quality due to the poor quality of the nominally focused image or/and due to the simultaneous presence of the defocused ghost image. The impact of ghosting may be further complicated because in most cases, contact lenses decenter when on the eye,[25] and move with the blink[26–27] and eye movements.[27] This alters the location of the defocused ghost image relative to the focused image,[21] and may affect visual quality. Also, multifocal lenses with spherical aberration that decenter on the eye will induce coma[6, 15, 28–29] that is in direct proportion to the amount of decentration and spherical aberration within the lens design.[30] Several other factors that may alter the appearance of the defocused (ghosted) image include: add power, zone geometry (number, arrangement, and size of zones), pupil size[31] and level of monochromatic aberrations in the eye.[20]

It is clear that ghost images may be a key factor in the success or failure of bifocal or multifocal presbyopic corrections, but we currently have no subjective method to quantify the multiple dimensions along which the ghosting may vary. In the present study we examine the ability of subjects to characterize four dimensions of ghost images when seen simultaneously with a well-focused image, and we examine the relationship between each of these perceptual dimensions and overall quality in a sample of 54 presbyopic patients fit with a multifocal contact lens.

MATERIALS AND METHODS

Part 1: Ghosting questionnaire development and validation

In the first part of this study, we evaluated the ability of subjects (n=10) to quantify four dimensions of ghosting using the printed score sheet shown in Figure 1a. Four ghosting dimensions were examined: direction, position-offset, focus/blur, and intensity. Each ghosted image was generated with custom computational optics software (MATLAB 4.2; Mathworks, Inc, Natwick, MA) and included both a focused letter and a simultaneously present ghosted letter “R,” which had its position-offset, direction, intensity and focus levels adjusted. The defocused letters were computed separately and added numerically to the focused image. On the printed score sheet the “direction series” of images consisted of a ghosted “R” being offset from a focused “R” in one of the 8 cardinal directions. In this series, the ghost portion of the letter had 0.50 D blur, a relative intensity of 50% of that of the focused image, and a 16 arc min offset. The “position-offset series” consisted of 10

separations, in which the ghosted "R" was offset horizontally from the focused "R" in increments of 1/10 of the letter width (equivalent to 3.2 arc min increments for these 20/160 letters). In this series, the direction of ghost offset was horizontal, the blur was 0.50 D, and the relative intensity was 50%. The "focus series" consisted of the ghosted "R" containing one of ten levels of focus ranging from 0.2–2.00D (computed for a pupil size of 3.5 mm and letter size of 20/160). Subjects scored how focused the ghost appeared. A score of 10 indicated a perfectly focused ghost, while a score of 1 indicated a ghost so blurred that no spatial detail could be identified. In this series, the ghost was offset by 8 arc min down and 8 arc min right, with a relative intensity of 50%. The "intensity series" consisted of 10 contrast levels spanning from a 100% contrast focused "R" and a 0% contrast ghosted "R" to a 100% contrast ghosted "R" and a 0% contrast focused "R". Letter (Weber) contrast increased in average (measured) steps of 11% per grade increment. In this series a focus level of 0.50 D was used, with an offset of 8 arc min down and 8 arc min right. In the offset, focus, and intensity dimensions, therefore, a higher score is associated with a more visible ghost.

A letter 'R' was chosen as a letter common to the Sloan letter set[32] that participants would be familiar with and would possess a range of characteristics common to most letters/symbols found in common text, including horizontal, vertical, radial, and curved lines. Steps of 10 were used for each attribute dimension. The number of steps was again chosen based on part on participant familiarity, as well as based on the number of steps required to provide adequate resolution throughout the entire desired range of each task. Previously, measurement reliability has been shown to steadily increase up to 7–10 steps, and then remain constant with increasing number of steps, so we did not feel the need for more than 10 steps.[33–34]

Ten pre-presbyopic subjects (6 male, 4 female) binocularly viewed single computationally ghosted images of black letter 'R' stimuli displayed on a linearized 26″ LED monitor (Apple, Cupertino, CA), with a white level luminance of 120cd/m^2. All subjects had best-corrected distance and near visual acuity of at least 20/20 and normal ocular health. Subjects wore their best sphero-cylindrical (distance) correction in a lightweight spectacle trial frame (UB-4 Universal Messbrille; Oculus; 12–15mm vertex distance) and viewed the display binocularly from 69 cm using a chin rest. Four series of ghosted images were displayed on the monitor: (1) ghosted 20/160 R images that exactly matched individual letters on the ghosting score sheet, (2) a series of ghosted 20/160 letters in which the ghosted letter was displaced in a different direction to those on the score sheet, (3) a series of 20/80 ghosted letters similar in all ways to those in the first series, in which the letter size and blur size were both reduced by a factor of two, and thus, retained the same relative scaling described above, (4) a series of ghosted 20/160 letters in which the four ghosting dimensions were randomly combined and thus, the displayed letters were not a perfect match to any single letter on the score sheet, but included direction, position-offsets, focus levels and intensities that were individually represented on the score sheet. Examples of each series are shown in Figure 1b. The subjects' task was to observe the single ghosted letter on the display and, using the printed score sheet, quantify the direction, position-offset, focus level and intensity of the ghost image. Note, in the case where all four ghosting dimensions varied simultaneously, is was impossible to distinguish a "ghost" and a "focused" image with intensity levels of 1 and 10 if these intensity levels occurred simultaneously with certain levels of focus and position-offset. As one could not even distinguish whether the letter observed was the ghost or focused letter, one was also unable to quantify any of the other ghosting dimensions. Therefore, in the current study, these two extreme intensity levels were omitted from the randomized trial sequences created prior to testing for the section of the experiment in which all dimensions were able to vary. These intensity levels were still included in testing when only one dimension varied at a time.

Before testing, sample images with no ghosting (e.g. only a focused "R" on the screen) and others containing visible ghost images in each of the 4 dimensions were used to explain the subject's task. During testing, subjects were allowed as much time as they needed to score the stimuli, but typically responses required around 10 seconds per presentation. Each stimulus level within each dimension was displayed in random order with 5 repetitions, resulting in 40–50 images per dimension.

A limits of agreement analysis[35] was used to assess the level of agreement between the measured and expected response as a function of response level. For each outcome we used mixed effects regression to model the association between the expected and observed values to allow for the correlation among repeated observations by each subject. We tested whether the slope and intercept of the regression line were zero. A p-value of 0.05 was used. Although we have identified seemingly independent ghosting dimensions, they may not segregate perceptually, and thus, subjects' reports of one dimension may be influenced by another dimension. Therefore, we used Analysis of Covariance (ANCOVA) to perform simultaneous comparisons between model-based estimates of the intercept and slope in the conditions where the stimulus and questionnaire directly matched (i.e. same size) relative to when the stimulus conditions did not directly match (i.e. different letter size, different orientation, etc). This procedure enabled us to perform a simultaneous test of both the intercepts and slopes between the matched and unmatched conditions. To do this analysis, we first created a dummy vector coded '0' for the reference condition and '1' for the comparison condition, and then regressed the difference scores (Observed-Expected) for the two conditions on a model that included a categorical main effect for condition, a continuous main effect for expected value, and the interaction between condition and expected value.

Linear mixed-effects modeling was used to investigate whether there were any systematic differences in the observed deviation scores (i.e., observed-expected) across the 5 trial repetitions (i.e., 1–5). Specifically, a separate model was evaluated for each stimulus condition using the data represented in Figures 2a – 4d. Each model included subject as a random effect and trial repetition both as a fixed effect and as a within-subject grouping factor. All models were fit by Restricted Maximum Likelihood (REML) estimation.

To guard against an inflation of the Type I error rate, follow-up pairwise comparisons were performed only if the mixed model yielded an effect of trial that was significant at $p<0.05$. Each comparison involved computing a t-test on the differenced deviation scores and evaluating the probability level for each 1-df contrast against the Bonferroni-adjusted rate of $0.05/10 = 0.005$. Note that the divisor for this per comparison rate incorporates all possible pairings of five trial means, i.e., $(5*(5-1))/2 = 10$.

Part 2: Clinical implementation of ghosting questionnaire

In the second part of the study, we implemented the validated ghosting questionnaire in a typical clinical situation to allow subjects wearing soft multifocal contact lenses to quantify the ghosting they perceived. In this part of the study fifty-four presbyopic subjects (43 females, 11 males) with a mean age of 56.3±4.5 years and a mean add of +2.00±0.50 Diopters wore a soft spherical and soft multifocal contact lens (MFCL) correction in one eye, while the other eye wore the best sphero-cylindrical near correction in a trial frame. All subjects had best-corrected distance and near visual acuities of at least 20/20 and normal ocular health. Each subject viewed a completely focused, high contrast (95%) 20/160 letter ('R') stimulus at 40cm with the MFCL corrected eye and viewed the ghosting questionnaire sheet at 40cm with their other eye wearing the best sphero-cylindrical near correction. Subjects observed the focused image with their MF eye, and scored any ghosting they observed along the four dimensions quantified on the ghosting score sheet used in part 1 of this study. Subjects also verbally rated the overall amount of ghosting using a standard 0–

100 numeric scale, where 0 was “Poor Quality - significant ghosting” and 100 was “Excellent Quality– “no ghosting”). The multifocal lenses used were a commercially available center near aspheric design (Air OptixAqua Multifocal, CIBA Vision, Duluth, GA). The spherical lenses were of the same material design, by the same manufacturer (Air Optix Aqua, CIBA Vision, Duluth, GA). All subjects had cylinder <0.75 D, and were fitted according to the manufacturer recommended fitting guide. Each subject wore the spherical monofocal contact lens design and one multifocal add design; the low, medium or high add. The average±std pupil size of the subjects was 4.5±1.3mm. The spherical aberration of these lens designs was measured using a validated contact lens aberrometer[36], because knowledge of the spherical aberration inherent in each lens designs is necessary as each design may be expected to contain slightly different levels of spherical aberration, which may then be expected to produce varying levels of quantifiable ghosting. Primary Zernike spherical aberration was quantified by the Z40 coefficients which, for all four types of contact lenses with a distance power of −6.00 D and a 6 mm pupil diameter, were −0.20, −0.30, −0.36, and −0.42 microns of Zernike (Z40) spherical aberration, for the monofocal, and three multifocal lenses, respectively. The mean sphere power (mean±std) of the contact lenses used in this study were 0.46±2.11 D, −1.38±1.49 D, −1.12±1.67 D, and −0.72±2.44 D, for the sphere, low, medium, and high add designs, respectively. Note that the sphere powers used were less minus as they included the necessary near add power. Additionally, based on a linear regression of subsequent contact lens aberrometry measures for 60 lenses of each design type across the power range of −8.00 D to +2.00 D, primary SA was found to vary by 0.02 microns per diopter for the sphere and medium add lenses, and by 0.01 microns per diopter for the low and high add designs. The four ghosting dimensions quantified with the scoring sheet used in the first part of this study were compared to their 0–100 numeric ratings of overall ghosting, using a multivariate regression analysis.

Ethical approval for this study was obtained from the Indiana University Institutional Review Board. All subjects gave written informed consent prior to beginning this study, and all procedures were conducted in accordance with the tenets of the Declaration of Helsinki.

RESULTS

Part 1: Ghosting questionnaire development and validation

The limits of agreement results (Figure 2–5) adopt a common graphical format. In each case, the diamond symbols represent the mean difference between the observed and expected (observed – expected) responses (y-axis) as a function of response level (x-axis) for each subject. The short-dashed lines represent the 95% limits of agreement for these difference scores, the solid line the mean difference (e.g., bias) between the responses, and the long-dashed lines note the best-fit linear regression of the difference scores. Note that when the stimulus level is 1, the difference scores can range from 0 to 9, when the stimulus level is 5, can range from −4 to +5, and when stimulus level is 10, the difference score can range from 0 to −9. Figures 2, 4, and 5 each contain four sub-panels describing the results of the (a) direction, (b) position-offset, (c) focus, and (d) intensity attributes for the “direct match” (e.g. stimulus presented on screen was identical to that on the score sheet)(Figure 2), “incomplete match” (e.g. stimulus presented on screen differed in one dimension from that on the score sheet)(Figure 4), and “completely non-matching” (e.g. no attribute was exactly the same in the stimulus presented on the screen as on the score card).

Overall, subjects were able to correctly match the presented stimuli to the correct value on the score sheet when there was a direct match in the attributes on the screen and score sheet (e.g. size (20/160) and position-offset (horizontally shifted ghost image)). Observed scores most closely matched the expected scores when matching the direction (Figure 2a), as neither the slope (m=0.00, p=0.71) nor intercept (b=−0.01, p=0.56) differed significantly

from 0, and the 95% LoA are less than 1/10 of the scoring increments (95% LoA: −0.08, 0.07). The position-offset judgments (Figure 2b) also closely matched the expected (m= −0.00, p= 0.70; b=−0.02, p=0.62), but with more variation in the responses generating wider 95% LoA, but still smaller than the scoring increments (95% LoA: −0.40, 0.33). Within the focus dimension, subjects had more difficulty directly matching the presented stimuli (Figure 2c). The 95% LoA were wider and approximately a single increment on the scoring sheet (95% LoA: −1.26, 0.87), and both the slope (m=−0.07, p<0.001) and intercept (b=0.17, p=0.02) were significantly different than 0. The significant negative slope indicates that on average, subjects reported that highly blurred targets (low focus score on the focus scale) appeared more focused or less blurred than expected, and, conversely, reported well-focused targets (e.g. ghosts with focus/blur levels of 8, 9 and 10) as less focused, or more blurred than expected. Performance on the intensity-matching (Figure 2d) task was quite similar to that of the focus matching (95% LoA: −0.85,0.74). Again, the slope (m=−0.04, p=0.005) was significantly different than 0. The intercept (b=0.16, p=0.063), however, was not significantly different than 0, indicating no average bias across the rating scale. Overall, however, the most significant result from this figure is that the ghosting attributes of direction, position-offset, focus and intensity can be accurately reported with errors very rarely exceeding 1 rating unit on the 8 and 10 point score scale used in this study.

Rating of ghosting position-offset performance when the 20/160 size stimuli were presented at an (a) oblique and (b) directly vertical position-offset are shown in Figure 3. These position-offsets were scored with the same score sheet that represented ghost offsets horizontally (Figure 1a). For both directions, the 95% LoA were slightly increased (95% LoA: −1.98, 1.21, and −0.97, 1.23, respectively) from the direct position-offset match described in Figure 2b, but again, errors were generally less than 1 step in the 10 point scale. In the oblique letter orientation (Figure 3a), the slope of the regression line of the difference scores was significantly different from 0 (p<0.001) and was also significantly different from that of the direct match (p<0.001) case shown in Figure 2b. There was no significant difference in intercept from 0 or from the direct match case. In the vertical deviation case (Figure 3b), the slope of the difference scores was not significantly different from 0 or from the direct match case (Figure 2b).

Overall performance when the subjects attempted to match the smaller (20/80) size stimuli presented on the screen with the larger (20/160) targets on the score sheet was quite similar to that of the 20/160 direct matching paradigm (Figure 4). Errors were generally less than 1 point on the scoring scale, the 95% LoA were virtually identical to the direct match case (direction 95% LoA: −0.08, 0.09; position-offset 95% LoA: −0.57, 0.49; focus 95% LoA: −1.32, 1.38; and intensity 95% LoA: −0.91, 0.77), and in all but the direction scoring, a small but statistically significant (p<0.05) negative slope to the data was observed.

As expected, variability in performance when all rating attributes were allowed to randomly vary (Figure 5) was higher than in either the direct match or orientation series. Specifically,, the 95% LoA were wider (direction 95% LoA: −1.27, 1.12; position-offset 95% LoA: −1.24, 1.99; focus 95% LoA: −1.90, 2.20; and intensity 95% LoA: −3.51, 1.40) relative to the perfect match cases (Figure 2), and the negative slopes were steeper (slope difference of −0.08, 0.03, −0.23, −0.38 units/D for the direction, position-offset, focus, and intensity attributes, respectively). Interestingly, the mean ghost intensity scores were lower than all other cases evaluated by 1.2 rating units. The negative slopes and positive intercepts indicated that on average, subjects significantly (p<0.01) over-estimated lower levels and under-score higher levels in every case, except for the position-offset attribute.

These results indicate that the direct matching task was easiest (Figure 2) and the completely non-matching combination of all randomly varying attributes (Figure 5) the most difficult.

However, in each condition, the majority of scores were within +/−1 rating unit of a perfect match (Figure 6). For example, in the direct matching task, participants were able to match all attributes within one rating unit on over 93% of the trials. In the case of the completely non-matching task, participants were still able to accurately describe the direction of the stimuli on 92% of the trials, position-offset on 84% of the trials, focus on 77% of the trials, and intensity on 53% of the trials.

When investigating the effect of trial repetition on the attribute ratings provided in the various conditions (i.e., do the results on 1 trial agree with the results of subsequent trials), mixed modeling analysis did not show any effect of trial repetition, except for the intensity series ($F_{4,36}$=2.878; p=0.036) (Figure 4d). Although follow-up t-tests using uncorrected probability levels suggest that paired comparisons involving the first trial differ significantly from all other trials, the Bonferroni-adjusted probability levels for these comparisons are all not significant. Overall, the modeling results indicate that there is no strong evidence of an effect of trial repetition in any of the stimulus conditions reported here. In other words, scores provided on the very first use of the questionnaire remain similar following repeated presentations, indicating little training may be required for the use of this questionnaire for short term evaluations like those employed in the current study.

Part 2: Clinical implementation of ghosting questionnaire

In the second experiment, which employed the real-world scenario of presbyopic subjects wearing MFCL utilizing the ghosting questionnaire validated in part 1 of this study, we examined how an overall numeric ghosting rating (0–100) co-varied with the scores of each of the four attributes included in the ghosting questionnaire (Figure 1). It is important to note that unlike in Part 1 of this experiment where presented images were always ghosted, and a corresponding ghosting rating was always needed, in Part 2 in some instances the subjects may not have noticed any ghosting, in which case they rated a 0 on all attribute scales. Additionally, the overall ghosting rating and ghosting questionnaire data sets employ reverse scales. In the individual attribute ghosting scales of the questionnaire (Figure 1), more visible ghosting (higher intensity ghost, better focused ghost, and more off-set ghosts) were assigned higher scores (e.g. 10), whereas the overall image ghosting rating scale employed 100 for no ghosting (high quality image without any sign of a ghost) and one for the case of very obvious ghosting. We scaled all of the individual attribute data to make it range from 0 to 100, and thus, expected slopes of −1 for scales that perfectly co-varied.

Figure 7 explores the covariance of each individual ghosting questionnaire attribute rating and the corresponding 0–100 overall ghosting score provided by the subject while wearing the same lens. A matrix of bivariate regression sub-panel scatterplots is shown in the upper off-diagonal sub-panels. The correlation data of interest are depicted at the “intersection” of the data listed in each row and column, respectively. For example, the first row and first column depict the correlation results for direction with each of the other attributes and overall 0–100 near rating. The second row and column depict the correlation of the position-offset attribute with all other attributes. The remaining panels depict the covariance of one attribute with another attribute. For example, the subpanel in the fourth column from the left and third row down from the top depicts the covariance between the focus and intensity attributes. In all cases, each sub-panel has: (1) the pairwise data grouped by lens type (i.e., sphere, low, medium, or high add), (2) 50 and 90% probability ellipses that correspond to 1.18 and 2.15 standard deviation units, respectively, and (3) a robust regression linear fit to all data points. Histograms representing the univariate response distribution of the ratings are shown along the diagonal sub-panels of the plot. Pairwise Spearman rank-order correlations are noted in the lower off-diagonal sub-panels, along with the corresponding estimated slope of the regression line in parentheses (i.e., bottom left square plot lists the correlation of ghosting direction and overall 0–100 near rating). As before, the data of

interest occur at the "intersection" of a corresponding row and column. For example, the correlation of the direction and intensity attributes is provided in the sub-panel in the first column from the left and fourth row from the top. As predicted, the correlations between individual attribute ratings and overall ghosting were significant ($p<0.001$) and negative in all cases. Also notice, however, that the 4 ghosting attributes are highly correlated with each other, with significant ($p<0.001$) correlations of around 0.90.

Several interesting results can be seen in Figure 7. First, when wearing the low, medium and high asphericity MFCLs, subjects reported a wide range of ghost focus levels (range 0–9), intensities (0–7), and ghost off-sets (0 – 6), in all possible directions of off-set. The latter two dimensions presumably related to the direction and magnitude of the contact lens decentration. However, subjects reported ghosting with any lens on only 35% (38/108) of the trials. Only twenty-three percent (11/54) of subjects wearing the sphere correction reported ghosting, whereas 33% (4/12) and 34% (8/23) reported ghosting with the low and medium add multifocal lens. Neither of these proportions of subjects reporting ghosting were significantly different than expected (one-sample proportions test with continuity correction, $p>0.5$ (i.e. 50%)). On the other hand, a significantly larger than expected proportion of subjects (79%; 15/19) wearing the high add lens reported ghosting.

Linear multiple regression analyses were then performed to assess the relative importance of the four stimulus-domain attributes on the observed overall numeric ghosting ratings. The initial model contained the response variable (i.e., ratings) and four explanatory variables only (i.e. direction, position-offset, focus, and intensity (no higher-order or interaction terms were included). The underlying rationale for excluding both non-linear and multiplicative terms from this model was that the 4 dimensional attributes (i.e., direction, position-offset, focus, and intensity) are, technically, uncorrelated in the stimulus domain. Based on all observations, the overall model was statistically significant ($F(4,104)=7.17$, $p<.0001$) and yielded a multiple R^2 of 0.22. The only significant explanatory variable in the model was due to the focus/blur dimension. The estimated coefficient for focus/blur was $b=-2.95$ ($t=2.32$, $p<0.05$), and was not dependent on the order in which variables were entered into the model. These modeling results indicate a substantial relative importance of focus/blur on rating judgments, and agree with the correlation results (Figure 7), in which the focus dimension had a correlation with each of the other dimensions of greater than 0.90.

However, given that a substantial proportion of the total variance in the ghosting ratings remained unexplained by the current model including focus/blur, a new model was formulated, which also contained the lens type worn during the rating (i.e. sphere, low, med, high add), as well as an interaction between focus and lens type. The overall test of this model was also statistically significant ($F(7,101)=5.97$, $p<.0001$) and yielded a multiple R^2 of 0.29. More importantly, the association between overall ghosting ratings and ghost focus was evident only for the Medium and High add multifocal designs, and was not present for the Spherical or Low add designs. This result highlights the fact that subjects may have experienced little perceived ghosting in the Spherical and Low add designs,

DISCUSSION

The first part of this study established that the four attributes of ghosting (direction, position-offset, focus/blur and intensity) can be quantified with high accuracy (90% of responses within +/− 1 of perfect using the 8 and 10 point visual scales on the ghosting score sheet). Ghost direction and position-offset were the easiest to correctly quantify, whereas the attributes of focus level and intensity were slightly more difficult. An error analysis revealed that confusion about which image was the ghost and which was the focused image was the

source of many errors. This most often occurred with medium intensity levels (e.g. ghost and focused image about the same contrast) and horizontal shifts.

Participants had most difficulty and overall porrest performance quantifying the intensity dimension which was also the most difficult dimension to accurately represent on the score sheet. Although the *average* contrast difference of the intensity series on the paper questionnaire was 11%, due to limitations in our photographic printing capabilities, the actual contrast step varied non-linearly from 8–14%, whereas the ghost intensities displayed on the monitor employed a fixed step of 11%. Therefore, these rendering errors between the score sheet and monitor were much smaller (<3%) than the step size, and would not be responsible for error scores greater than 1.

In evaluating the focus/blur dimension in more detail, it appeared that subjects were hesitant to utilize the extreme ends of the scale. It is uncertain why this occurred, but could merely be that subjects also wanted to leave available ratings for the image to get “better” or “worse”. It is clear, however, that as the complexity of the task increased (e.g. not directly matching, smaller target sizes, more than one dimension to consider at a time), accuracy decreased.

In many cases, a fit to the difference scores yielded a significant negative slope, potentially indicating that on average, subjects overestimate their rating at lower rating levels, and underestimate their ratings at higher levels of blur. However, given that a participant cannot give a score under 1 or above 10, a line of negative slope may actually be the expectation when employing this analysis of data of this type. To investigate this theory, we refit the data shown in Figures 2–5 restricting the range of our analysis to ratings units of 3–8 only, and then evaluated the significance of the slope of these regression lines. For the direct match case (Figure 2), the varied direction case (Figure 3), and the smaller letter size case (Figure 5), all slopes that were previously significant became no longer significant. However, in the completely non-matching case (Figure 5), where variance was greatest, the slopes remained significantly different than 0 ($p<0.01$).

Overall, however, the clinical implementation of the questionnaire demonstrates its general agreement with the standard numeric rating scale currently typically used to quantify ghosting [8]. There was a strong association between the overall numeric representation of ghosting and each ghosting attribute explored, indicating that overall perceived ghosting levels increased as ghost intensity, ghost off-set, and ghost focus increased. Conversely, less ghosting was perceived with ghosted images shifted less, blurred more, or when the ghost had lower intensity. Of the four attributes explored, however, “focus/blur” was the main driver of ghosting numeric ratings, with higher levels of blur in the ghosted image associated with better overall numeric ratings. This indicates that if the ghosted image is blurred to a certain level, the noticeability of the ghosted image is decreased. If, however, the ghost is well-focused, subjects will report more ghosting. Ghost position-offset was also significantly correlated with overall perceived ghosting and therefore, as add power increases, and blur of the ghost will generally increase, Prentice’s Rule predicts that ghost offset will increase. Thus, the potential strategy of reducing ghosting by employing higher add powers may fail in the presence of significant lens decentration.

Although this ghosting questionnaire had been employed briefly in a pilot setting, this was our first validation and implementation of a scale of this nature. As a result, there are several issues with the current scale that may need to be refined. First, as highlighted above, in the case of intensity ratings of 1 and 10, combined with low levels of defocus (high focus rating) and position-offset, it is impossible to distinguish whether the dominant aspect of the stimulus is the ghost or focus portion of the image. Fortunately, however, it is believed this

combination of attributes (low contrast focused image and high contrast ghosted image) would not be one that would commonly occur in the clinical setting. Second, also as highlighted above, the variation in the three lowest focus ratings (highest amounts of dioptric blur) were quite high, indicting an inability of subjects to distinguish between these blur levels. This indicates that refined levels of focus/blur with greater separation between levels may be needed within a scale. Third, as the description and reference of the terminology becomes difficult to describe and remember at times (e.g. low focus rating corresponding to large amounts of defocus), some revision in the scale terminology and ordering may ultimately make it easier to clinically implement. Lastly, the current ghosting questionnaire approach only allows monocular quantification, whereas clinically someone may most commonly wear a design that may cause ghosting in both eyes. Previous literature suggests that the presence of ghosted images in each eye simultaneously may cause anomalous fusion leading to the perception of a "depth" disparity or "3-D" effect in some cases [37–38] However, further work is needed to investigate this effect in more depth.

Regardless of these potential issues, the novel method developed to quantify monocular levels of perceived ghosting provides accurate assessments of the perceived ghosted image along four stimulus dimensions, and this technique can be easily employed in the clinical environment to evaluate ghosting with different lens designs.

Acknowledgments

Support of BoKaye Dietmeier by T35EY013937.

References

1. Partal AE, Manche EE. CustomVue laser in situ keratomileusis for myopia and myopic astigmatism using the Visx S4 excimer laser: Efficacy, predictability, and safety. J Cataract Refract Surg. 2006; 32(3):475–9. [PubMed: 16631061]
2. Durrie DS, Stahl J. Randomized comparison of custom laser in situ keratomileusis with the Alcon CustomCornea and the Bausch & Lomb Zyoptix systems: one-month results. J Refract Surg. 2004; 20(5):S614–8. [PubMed: 15523985]
3. Michalewska Z, Michalewski J, Odrobina D, Nawrocki J. Non-Full-Thickness Macular Holes Reassessed with Spectral Domain Optical Coherence Tomography. Retina. 2011
4. Querques G, Querques L, Rafaeli O, Canoui-Poitrine F, Bandello F, Souied EH. Preferential hyperacuity perimeter as a functional tool for monitoring exudative age-related macular degeneration in patients treated by intravitreal ranibizumab. Invest Ophthalmol Vis Sci. 2011; 52(9):7012–8. [PubMed: 21885622]
5. Back A, Grant T, Hine N, Holden BA. Twelve-month success rates with a hydrogel diffractive bifocal contact lens. Optom Vis Sci. 1992; 69(12):941–7. [PubMed: 1300517]
6. Patel S, Fakhry M, Alio JL. Objective assessment of aberrations induced by multifocal contact lenses in vivo. Clao J. 2002; 28(4):196–201. [PubMed: 12394546]
7. Martin JA, Roorda A. Predicting and assessing visual performance with multizone bifocal contact lenses. Optom Vis Sci. 2003; 80(12):812–9. [PubMed: 14688544]
8. Papas EB, Decenzo-Verbeten T, Fonn D, et al. Utility of short-term evaluation of presbyopic contact lens performance. Eye Contact Lens. 2009; 35(3):144–8. [PubMed: 19421021]
9. Applegate RA, Marsack JD, Ramos R, Sarver EJ. Interaction between aberrations to improve or reduce visual performance. J Cataract Refract Surg. 2003; 29(8):1487–95. [PubMed: 12954294]
10. West SK, Rubin GS, Broman AT, Munoz B, Bandeen-Roche K, Turano K. How does visual impairment affect performance on tasks of everyday life? The SEE Project. Salisbury Eye Evaluation. Arch Ophthalmol. 2002; 120(6):774–80. [PubMed: 12049583]
11. Carta A, Braccio L, Belpoliti M, et al. Self-assessment of the quality of vision: association of questionnaire score with objective clinical tests. Curr Eye Res. 1998; 17(5):506–11. [PubMed: 9617546]

12. Maldonado-Codina C, Efron N. Impact of manufacturing technology and material composition on the clinical performance of hydrogel lenses. Optom Vis Sci. 2004; 81(6):442–54. [PubMed: 15201718]
13. Chu BS, Wood JM, Collins MJ. Effect of presbyopic vision corrections on perceptions of driving difficulty. Eye Contact Lens. 2009; 35(3):133–43. [PubMed: 19421020]
14. Papas EB, Schultz BL. Repeatability and comparison of visual analogue and numerical rating scales in the assessment of visual quality. Ophthalmic Physiol Opt. 1997; 17(6):492–8. [PubMed: 9666923]
15. Kollbaum PS, Jansen ME. Comparison of patient-reported visual outcome methods to quantify the perceptual effects of defocus. Contact Lens and Anterior Eye. 2011
16. Ravikumar S, Bradley A, Thibos L. Phase changes induced by optical aberrations degrade letter and face acuity. Journal of Vision. 2010; 10(14)
17. Yellott JIYJ. Correcting spurious resolution in defocused images. Proceedings of SPIE. 2007:6492.
18. Verhoeff FH. The cause of a special form of monocular diplopia. Arch Ophthalmol. 1900; 29:565–572.
19. Hirohara Y, Mihashi T, Suzuki A, Kuroda T. Evaluating optical quality of a bifocal soft contact lens in near vision using a Shack-Hartmann wavefront aberrometer. Optical Review. 2006; 13(5): 396–404.
20. Iskander DR, Collins MJ, Davis B, Carney LG. Monochromatic aberrations and characteristics of retinal image quality. Clin Exp Optom. 2000; 83(6):315–322. [PubMed: 12472423]
21. Charman WN, Walsh G. Retinal Image Quality with Different Designs of Bifocal Contact Lens. Trans BCLA Conference. 1986:13–19.
22. Hutnik CM, O'Hagan D. Multifocal contact lenses--look again! Can J Ophthalmol. 1997; 32(3): 201–5. [PubMed: 9131286]
23. Negishi K, Ohnuma K, Ikeda T, Noda T. Visual simulation of retinal images through a decentered monofocal and a refractive multifocal intraocular lens. Jpn J Ophthalmol. 2005; 49(4):281–6. [PubMed: 16075326]
24. Terwee T, Weeber H, van der Mooren M, Piers P. Visualization of the retinal image in an eye model with spherical and aspheric, diffractive, and refractive multifocal intraocular lenses. J Refract Surg. 2008; 24(3):223–32. [PubMed: 18416256]
25. Tomlinson A, Bibby MM. Movement and rotation of soft contact lenses. Effect of fit and lens design. Am J Optom Physiol Opt. 1980; 57(5):275–9. [PubMed: 7386591]
26. Tomlinson A, Ridder WH 3rd, Watanabe R. Blink-induced variations in visual performance with toric soft contact lenses. Optom Vis Sci. 1994; 71(9):545–9. [PubMed: 7816424]
27. Young G. Evaluation of soft contact lens fitting characteristics. Optom Vis Sci. 1996; 73(4):247–54. [PubMed: 8728492]
28. Peyre C, Fumery L, Gatinel D. Comparison of high-order optical aberrations induced by different multifocal contact lens geometries. J Fr Ophtalmol. 2005; 28(6):599–604. [PubMed: 16141922]
29. Gatti RF, Lipener C. Optical performance of different soft contact lenses based on wavefront analysis. Arq Bras Oftalmol. 2008; 71(6 Suppl):42–6. [PubMed: 19274410]
30. Guirao A, Williams DR, Cox IG. Effect of rotation and translation on the expected benefit of an ideal method to correct the eye's higher-order aberrations. J Opt Soc Am A Opt Image Sci Vis. 2001; 18(5):1003–15. [PubMed: 11336203]
31. Charman WN, Jennings JA, Whitefoot H. The refraction of the eye in the relation to spherical aberration and pupil size. Br J Physiol Opt. 1978; 32:78–93. [PubMed: 737383]
32. Sloan LL. New test charts for the measurement of visual acuity at far and near distances. Am J Ophthalmol. 1959; 48:807–13. [PubMed: 13831682]
33. Streiner, DL.; Norman, GR. Health measurement scales: a practical guide to their development and use. 2. Oxford ; New York: Oxford University Press; 1995.
34. Hand, DJ. Measurement theory and practice: the world through quantification. London New York: Arnold ; Distributed in the U.S.A. by Oxford University Press; 2004.
35. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986; 1(8476):307–10. [PubMed: 2868172]

36. Kollbaum P, Jansen M, Thibos L, Bradley A. Validation of an off-eye contact lens Shack-Hartmann wavefront aberrometer. Optom Vis Sci. 2008; 85(9):E817–28. [PubMed: 18772713]

37. McGill E, Erickson P. Stereopsis in presbyopes wearing monovision and simultaneous vision bifocal contact lenses. Am J Optom Physiol Opt. 1988; 65(8):619–26. [PubMed: 3177586]

38. Holley B, Carmichael C. The Alges Bifocal Contact Lens. Contact Lens Forum. 1987; 2:52–4.

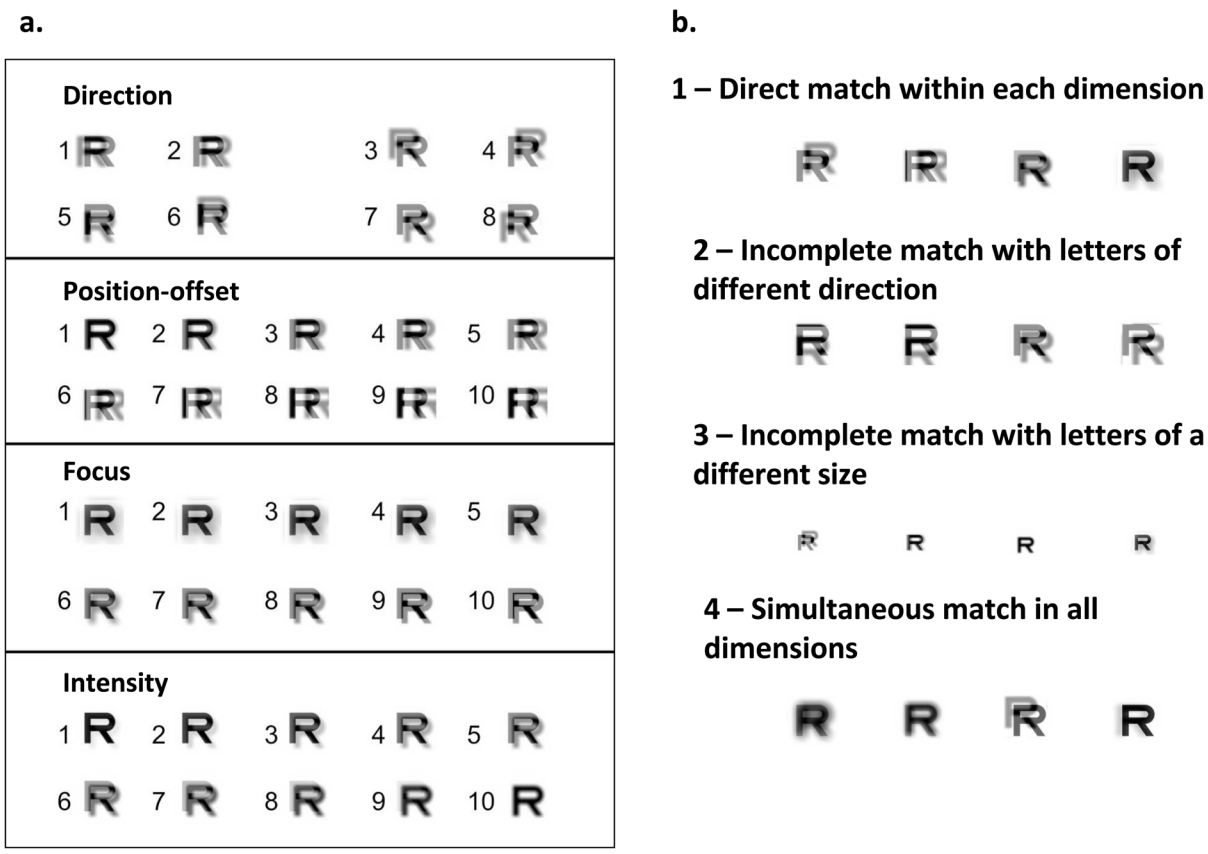

Figure 1.
(a.) Printed Questionnaire (FUJIFILM Fujicolor Crystal Archive photographic paper) consisting of four image dimensions series: Direction (8 images), Position-offset (10 images), Focus (10 images) and Intensity (10 images). (b.) Sample images for each of the 4 series of ghosted images.

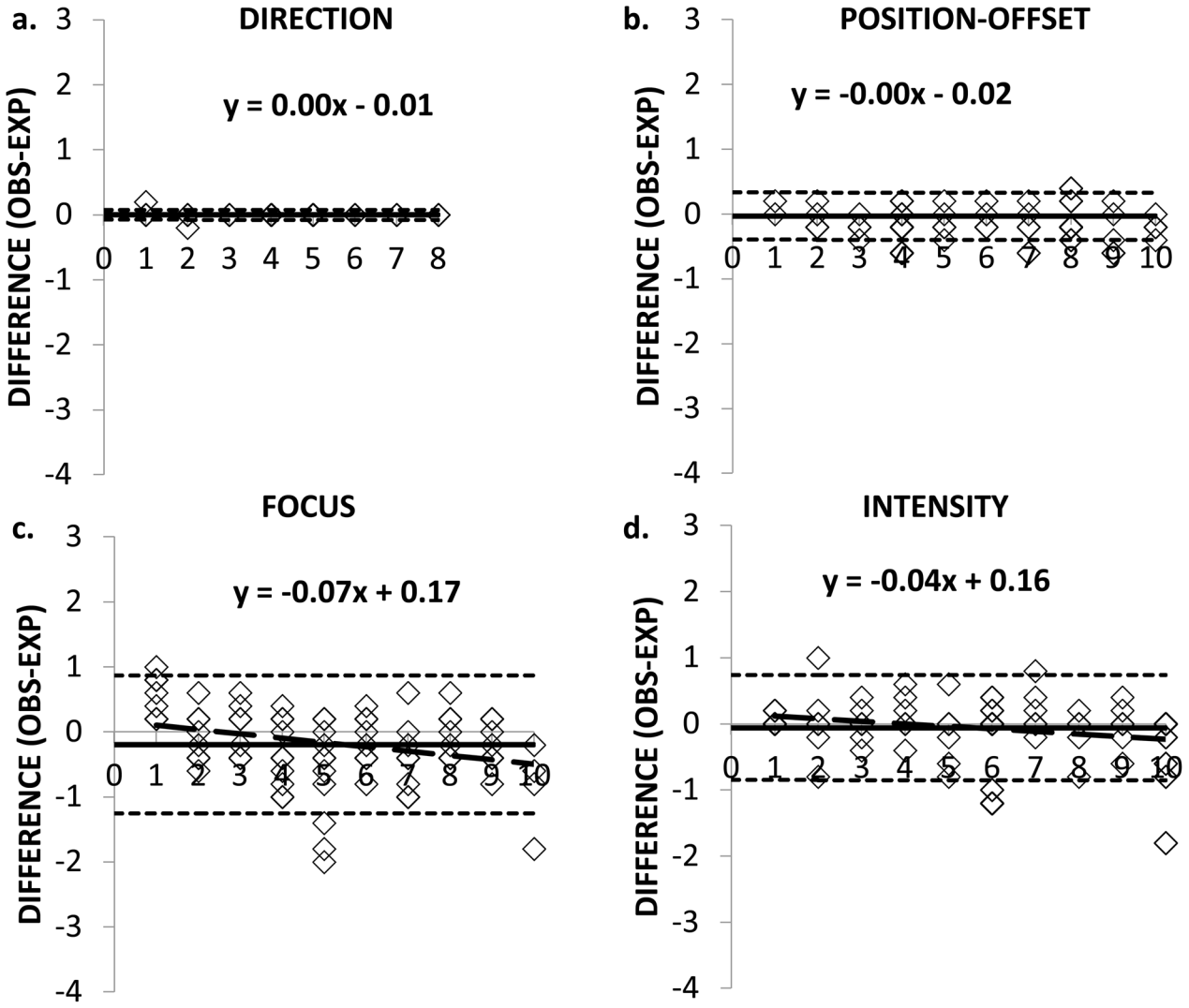

Figure 2.

Limits of agreement for the difference between the observed and expected rating responses for the ghosting attributes of (a.) Direction, (b.) Position-offset, (c.) Focus, and (d.) Intensity for the direct match case of a 20/160 target on the monitor and this same size target on the ghosting questionnaire. The symbols represent the mean difference for each subject's five trials for each rating number in the series. The x-axis represents the expected rating on the ghosting questionnaire. The solid black horizontal line represents the mean difference score for all subjects and rating levels. The black long dashed line represents the best-fit linear regression line, and the short dashed black horizontal lines the 95% Limits of Agreement for the difference scores. Note that when the stimulus level is 1, the difference scores can range from 0 to 9, when the stimulus level is 5, can range from −4 to +5, and when stimulus level is 10, the difference score can range from 0 to −9.

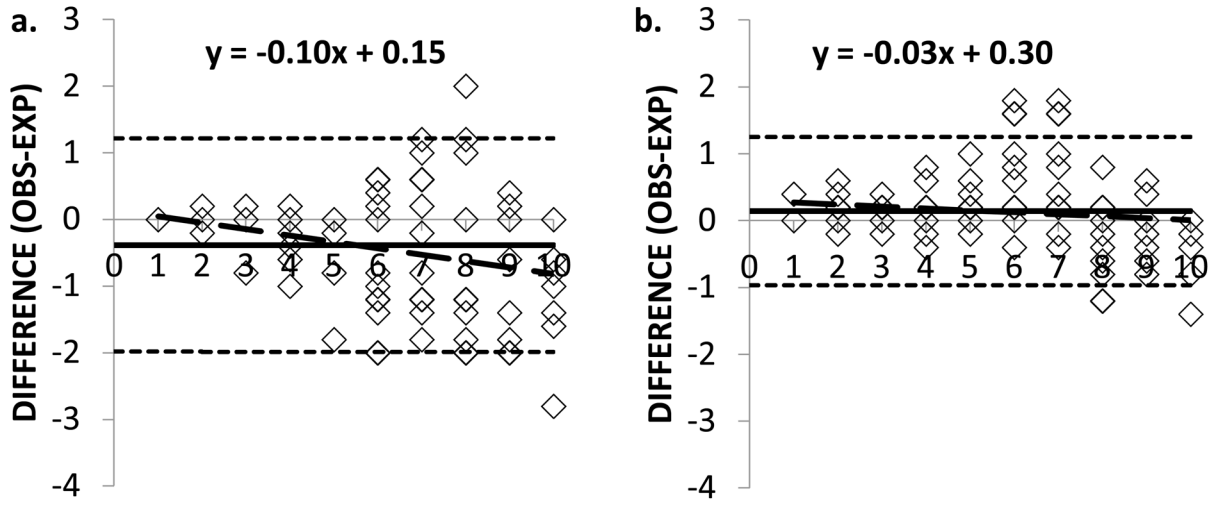

Figure 3.

Limits of agreement for the difference between the observed and expected rating responses for the ghosting attributes of an (a.) Oblique and (b.) Vertical ghosted image position-offset for a 20/160 target. The symbols represent the mean difference for each subject's five trials for each rating number in the series. The x-axis represents the expected rating on the ghosting questionnaire. The solid black horizontal line represents the mean difference score for all subjects and rating levels. The black long dashed line represents the best-fit linear regression line, and the short dashed black horizontal lines the 95% Limits of Agreement for the difference scores. Note that when the stimulus level is 1, the difference scores can range from 0 to 9, when the stimulus level is 5, can range from −4 to +5, and when stimulus level is 10, the difference score can range from 0 to −9.

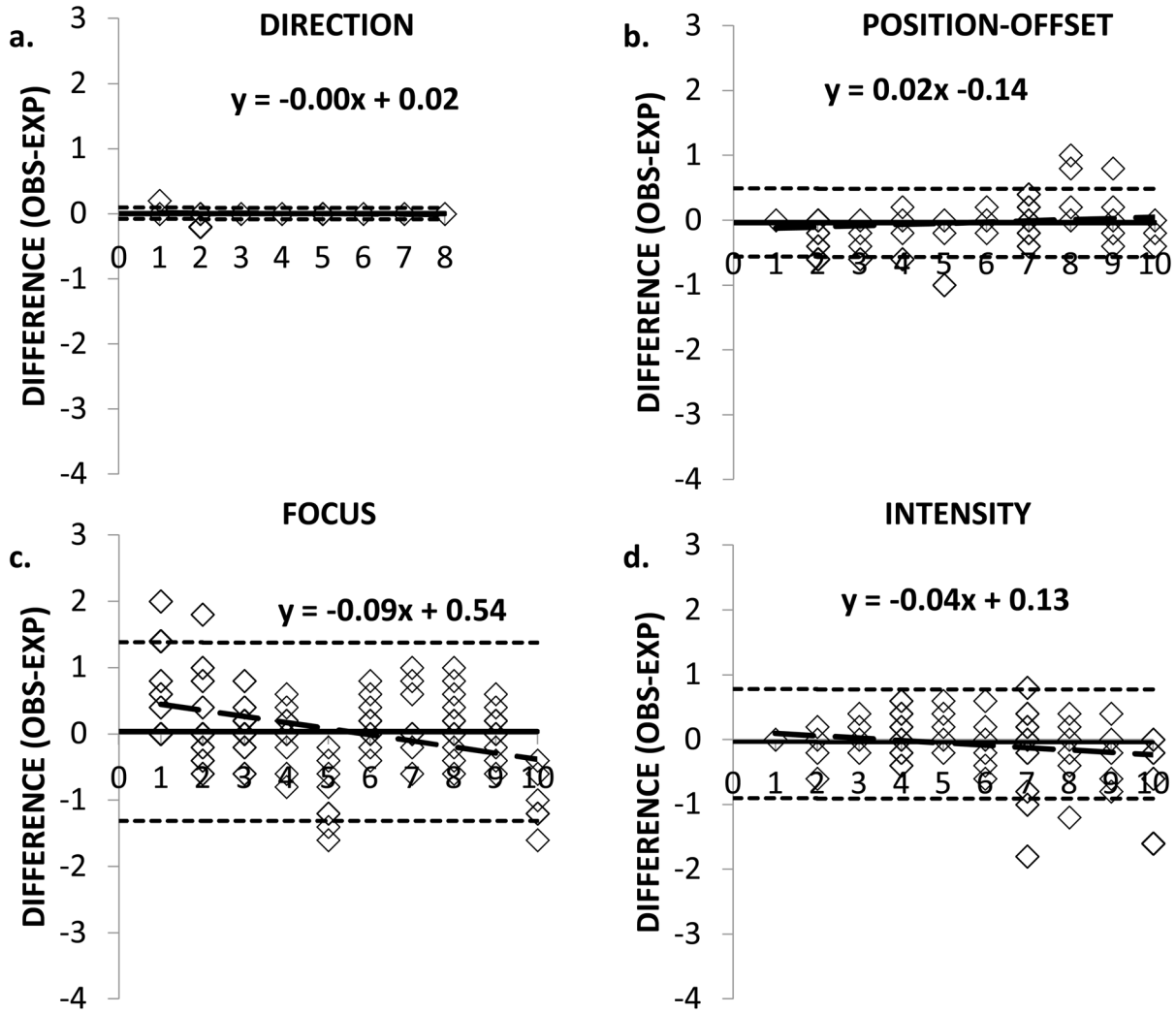

Figure 4.

Limits of agreement for the difference between the observed and expected rating responses for the ghosting attributes of (a.) Direction, (b.) Position-offset, (c.) Focus, and (d.) Intensity for a 20/80 size stimulus. The symbols represent the mean difference for each subject's five trials for each rating number in the series. The x-axis represents the expected rating on the ghosting questionnaire. The solid black horizontal line represents the mean difference score for all subjects and rating levels. The black long dashed line represents the best-fit linear regression line, and the short dashed black horizontal lines the 95% Limits of Agreement for the difference scores. Note that when the stimulus level is 1, the difference scores can range from 0 to 9, when the stimulus level is 5, can range from −4 to +5, and when stimulus level is 10, the difference score can range from 0 to −9.

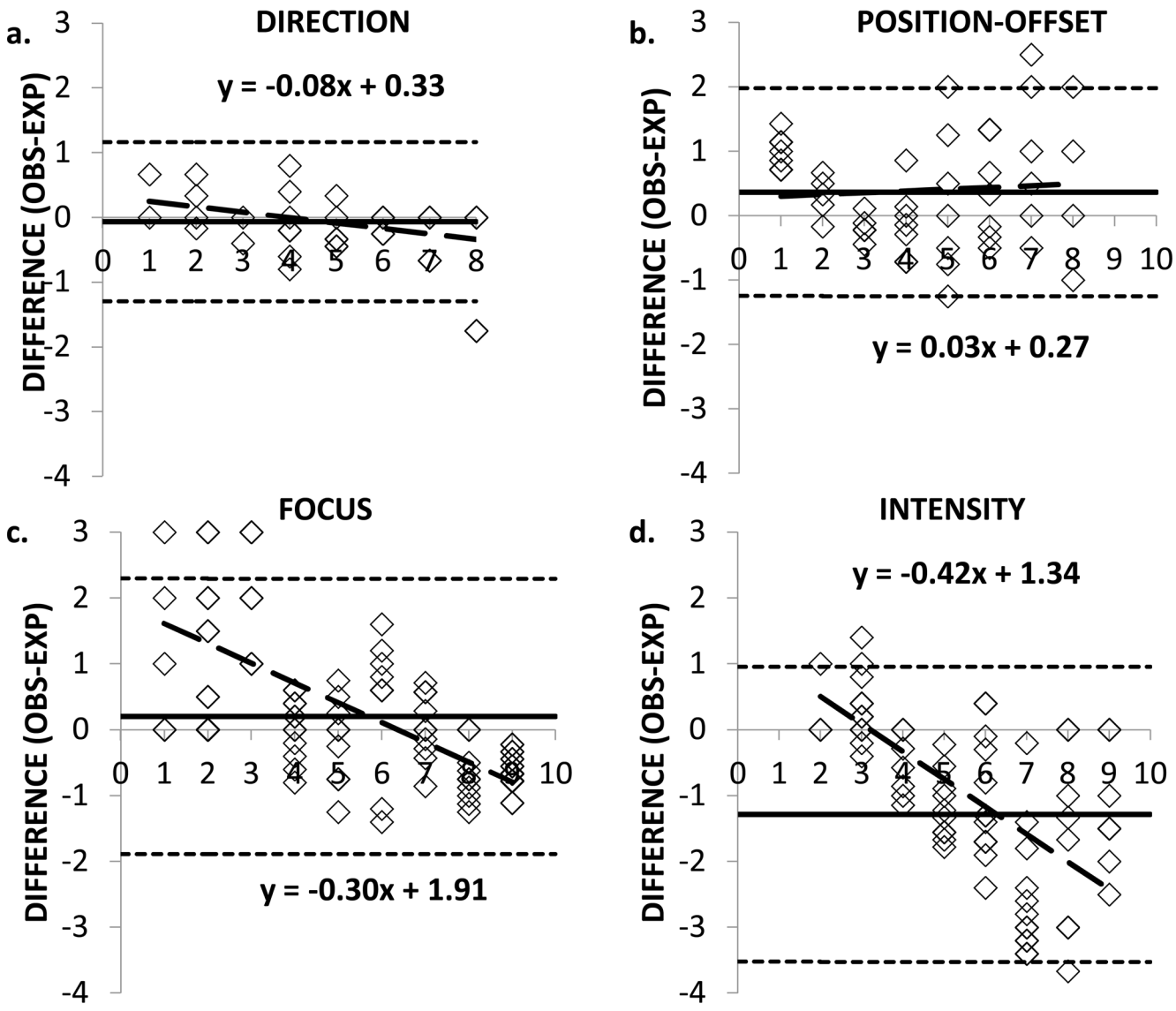

Figure 5.

Limits of agreement for the difference between the observed and expected responses for the simultaneous presentation of randomized attribute combinations for the following ghosting attributes of (a.) Direction, (b.) Position-offset, (c.) Focus, and (d.) Intensity for a 20/160 target. The symbols represent the mean difference for each subject's five trials for each rating number in the series. The x-axis represents the expected rating on the ghosting questionnaire. The solid black horizontal line represents the mean difference score for all subjects and rating levels. The black long dashed line represents the best-fit linear regression line, and the short dashed black horizontal lines the 95% Limits of Agreement for the difference scores. Note that when the stimulus level is 1, the difference scores can range from 0 to 9, when the stimulus level is 5, can range from −4 to +5, and when stimulus level is 10, the difference score can range from 0 to −9.

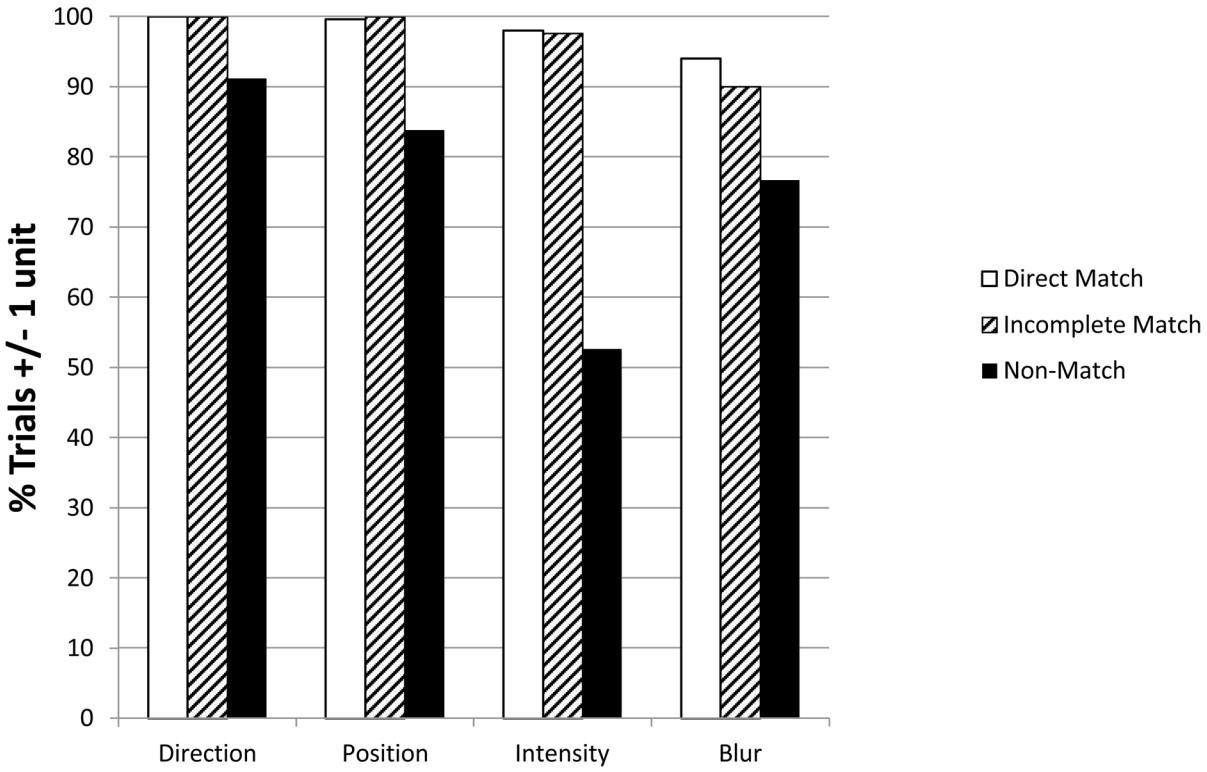

Figure 6.

Grouped barplots for each ghosting attribute depicting the percentage of trials that were within +/−1 unit for the Direct Match series (i.e., matching 20/160 images of the questionnaire to 20/160 targets on the monitor)(white bars), the Incomplete Match series (i.e., matching 20/80 images on the monitor to the 20/160 images on the questionnaire, but other stimulus aspects remained the same)(hatched bars), and the Non-Match series (i.e., simultaneous presentations of random combinations of all stimulus dimensions of 20/160 letter sizes)(black bars).

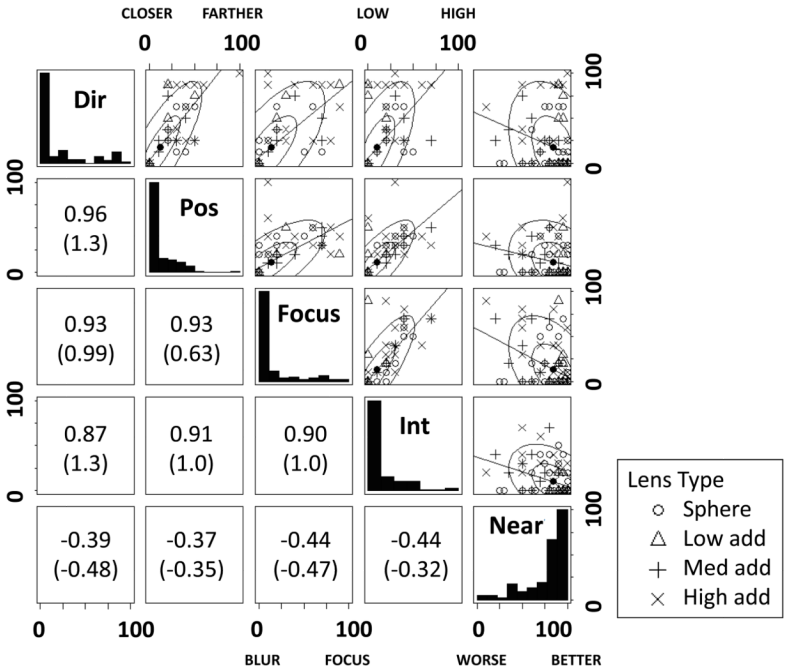

Figure 7.

A matrix of bivariate regression scatterplots is shown in the upper off-diagonal panels. Each panel has: (1) the pairwise data grouped by lens type (i.e. sphere, low, medium, or high add), (2) 50 and 90% probability ellipses that correspond to 1.18 and 2.15 standard deviation units, respectively, and (3) a robust regression linear fit to all data points. Histograms representing the univariate distribution of each response are shown along the diagonal panels of the plot. Pairwise Spearman rank-order correlations are noted in the lower off-diagonal cells, along with the estimated slope of the regression line in parentheses. The first row and first column depict the correlation results for direction with each of the other attributes and overall 0–100 near rating. The second row and column depict the correlation of the position-offset attribute with all other attributes. The third row and third column depict the results of the focus/blur dimension, and the fourth row and column depict the results of the intensity dimension.