

Recognition performance and perceived quality of video enhanced for the visually impaired

Eli Peli

The Schepens Eye Research Institute, Harvard Medical School, 20 Staniford Street, Boston, MA 02114-2500, USA

Abstract

Subjects with central field loss (CFL) individually selected enhancement parameters to improve visibility of static video images. The effect of enhancement on performance and on perceived quality of motion video was assessed. Performance, e.g. recognition of visual details, was assessed by having subjects answer questions regarding visual information contained in the video motion segments that were enhanced using the individually-selected parameters. Enhancement did not improve subject performance on questions about video content. This result might be due to a ceiling effect limitation of the performance assessment method. In a second procedure, subjects' continuous perceptions of quality (using an adjective-based rating scale) were made while the enhancement parameters were abruptly switched among multiple values; these included the individually-selected enhancements, as well as unenhanced, over-enhanced, and degraded segments. The results indicate that adaptive enhancement (individually-tuned using a static image) adds significantly to perceived image quality when viewing motion video. Subjects who selected stronger contrast enhancement also perceived the enhancement to provide a larger benefit in image quality.

Keywords: image enhancement, low vision, perceived quality, television, video, vision rehabilitation

Introduction

Enhancement of video images has been suggested as an aid for the visually impaired (Peli and Peli, 1984; Peli *et al.*, 1986; Isenberg *et al.*, 1989). The basic approach proposed by Peli and adopted by others for enhancement of images (Isenberg *et al.*, 1989; Myers *et al.*, 1995; Omoruyi *et al.*, 2001), as well as for the enhancement of displayed text (Lawton, 1992; Fine and Peli, 1995) was to use (high) band pass filtering to enhance those spatial frequencies that are not detected by the visually impaired at their naturally occurring low contrast (Peli *et al.*, 1991; Peli *et al.*, 1994b). Peli *et al.* have investigated a few modifications of this contrast enhancement

approach; applying wide-band enhancement by superposition of high contrast outlines over the images (Peli *et al.*, 2004) as well as band pass enhancement implemented in the MPEG compression domain (Kim *et al.*, 2004).

A different approach to image processing for the visually impaired (frequently called image enhancement in the literature) involves the segmentation and classification of image regions and marking relevant image segments with either high contrast or different colour or any other (possibly non-visual) markers (Everingham *et al.*, 2003; Tu *et al.*, 2003; Zur and Ullman, 2003; Bryant *et al.*, 2004a,b). This segmentation and classification approach is most commonly envisioned to be a part of a mobility aid, while the enhancement through spatial filtering reported here is intended mostly for TV viewing.

One form of enhancement [Adaptive Enhancement (Peli and Lim, 1982)] has been demonstrated to be effective in optical simulations (Peli and Peli, 1984), in computational simulations of static images (Peli *et al.*, 1991), and in improving face recognition from static

Received: 27 August 2004
Revised form: 18 March 2005
Accepted: 16 April 2005

Correspondence and reprint requests to: Eli Peli.
Tel.: +1 617 912-2597; Facsimile: +1 617 912-0112.
E-mail address: eli@vision.eri.harvard.edu

images by subjects with central scotoma and by subjects with cataracts (Peli *et al.*, 1991). For many of these subjects the effect of the uniformly applied enhancement was found to be significant and for a number of them, the magnitude of the improvement was large.

All the image enhancement work cited above was carried out with static images, except for Kim *et al.*'s (2004) study that used short video segments pre-processed at different levels and replayed from disk for the subjects. The development of the DigiVision CE-3000 device (Hier *et al.*, 1993) made it possible to apply the adaptive enhancement algorithm in real time to motion video with real-time modifications to the enhancement parameters. In a pilot study (Peli *et al.*, 1994a), visually impaired subjects viewed a frozen video image and were asked to use the CE-3000 control knobs to adjust the picture until they could see as many details as possible. The subjects were then shown a motion-video segment, both with and without the individually-selected enhancement, asked for a preference between the two presentations, and were further asked to make specific judgments about the effect of the processing on image appearance measures. This was followed by questions about a few visual details contained in several movie scenes viewed both in the enhanced and unenhanced modes. The subjects were able to distinguish between the unenhanced and enhanced images. All preferred the enhanced version of the still images to the unenhanced versions, and 95% preferred the enhanced motion segments. During informal interviews, these subjects stated that the enhancement resulted in darker images with more apparent outlines and a greater number of fine details than the original videos. Surprisingly, 90% thought the enhanced image appeared more natural. There was also a statistically significant, though small, improvement in recognition of details in scenes taken from a motion picture.

The enhancement in that pilot study was individually selected for each subject by turning the two CE-3000 knobs controlling the Detail and Contrast levels (Background knob was fixed) in a free search procedure, which many subjects found to be a very difficult task. In a later study, using a fixed enhancement applied for all subjects, only 21% of the subjects indicated a preference for the enhancement (Fine *et al.*, 1997). In both studies, the preference was evaluated by questions presented following sequential viewing of both the enhanced and unenhanced versions of the same video segment.

The studies reported in this paper were conducted to further assess the effect of individually-selected enhancements on both recognition performance and perceived image quality during the viewing of motion video segments. The first procedure consisted of measuring *recognition performance* using an 'Audio Description' (AD) based questionnaire (Peli *et al.*, 1996). A subset of

these subjects, in a separate procedure, provided *continuous ratings of perceived quality* for video segments that were enhanced using individually-selected enhancement parameters as well as other enhancement settings. The subjects' rating of the video quality was continuously recorded in response to modifications of enhancement parameters. In addition to the *individually-selected enhancement parameters*, arbitrary modifications of the enhancement parameters (applied in two different schemes – see below) including image degradation and over-enhancement conditions were tested. The perceived quality was compared with that of the original unenhanced videos.

Methods

General

Four different procedures were used in this study: (1) The individual selection of enhancement parameters using static images; (2) The evaluation of the effect of those enhancements on recognition performance using the AD-based questionnaire; (3) A comparison judgment of overall quality of enhanced and unenhanced video segments shown during procedure (2); and (4) A continuous evaluation of perceived image quality while viewing motion video segments processed with varying enhancement parameters. Each of these procedures is described in detail below.

Subjects viewed the screen with their habitual optical correction used for TV viewing at home. Sitting distance from the screen was adjusted to match the angular substance to that experienced by the subjects viewing their own TV at home.

Apparatus

Subjects viewed the video played through a Sony Umatic SP VO-9600, a professional grade 3/4" video-cassette recorder (controlled by a Macintosh computer), on a 27" Sony Trinitron colour television monitor. The Public Television station WGBH in Boston developed the AD for these programs and provided the program videotapes. Programs were on high quality 3/4" tapes providing broadcast quality. Both the videotape recorder and the monitor were at the default factory setting for contrast and luminance. Image enhancement was accomplished through a specially modified DigiVision CE-3000 device (DigiVision Inc., San Diego, CA, USA) that implements, in real time (Hier *et al.*, 1993), the adaptive enhancement algorithm (Peli and Peli, 1984). The video signal is separated into luminance (brightness) and chroma (colour) components, and the enhancement is applied to the luminance component only. Examples of the

effects of this processing are shown in *Figure 1*. The parameters of the algorithm are controlled on the CE-3000 via three independent knob settings labelled *detail*, *contrast*, and *background*. The relations of these settings to the parameters of the adaptive enhancement algorithm (Peli and Peli, 1984) are described below. On the standard device all controls have 10 possible analogue settings, ranging from 0 to 9. In addition to the manual parameter settings, the modified device enabled computer control, via a graphics tablet, of 100 settings (0–99) for each parameter.

The *detail* control determines the S.D. of a Gaussian-shaped averaging window. This window is used to compute a local mean luminance (low-pass filtered) image that is then subtracted from the original image. The residual (high-pass filtered) image is multiplied by the *contrast* parameter to provide the enhancement. Higher detail values result in a smaller averaging window and as a result in the enhancement of higher spatial frequencies. A detail setting of 99 corresponds to an averaging window in the adaptive enhancement algorithm (Peli and Lim, 1982) with a Gaussian sigma of 2% of image size (resulting in enhancement of high spatial frequency details – of the order of 2% of image size), while the detail setting of 0 corresponds to an averaging window with a sigma of 16% of image size (resulting in filtering starting at a fairly low spatial frequency – above about six cycles per image). The up/down (or near/far) direction of the graphics tablet was used to control this detail parameter with higher detail values generated by mouse positions farther away from the subject.

The *contrast* parameter amplifies the high-pass filtered image. This is equivalent to the K parameter in Peli and

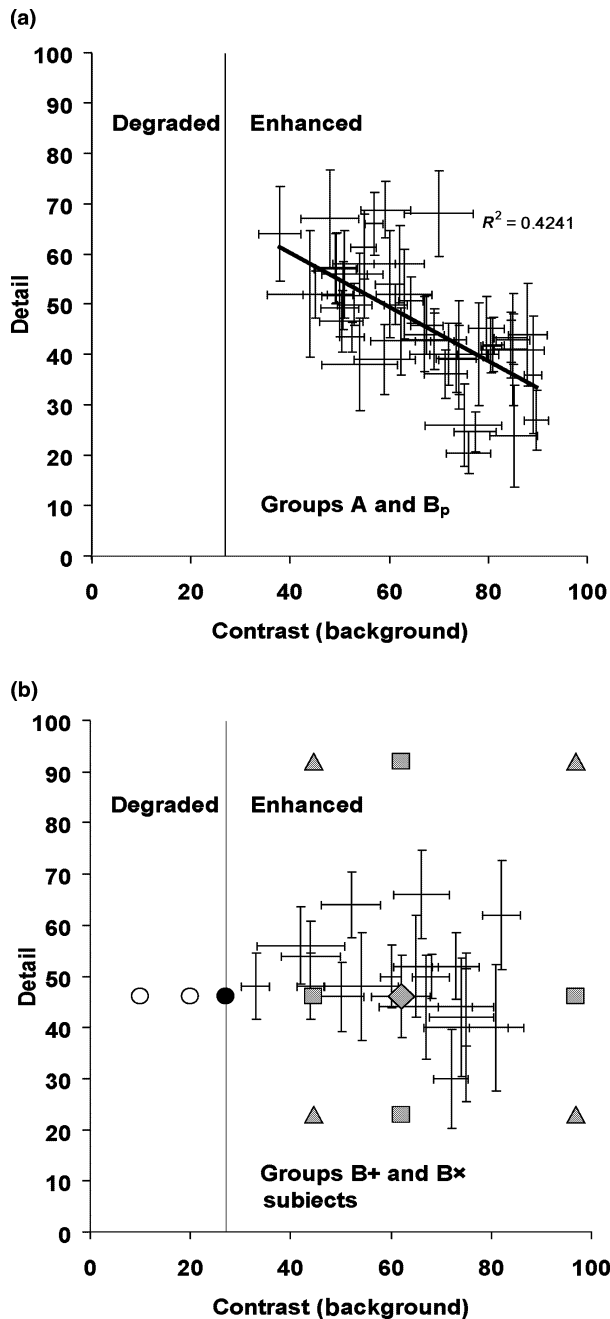
Peli (1984). For the purpose of the current study, the DigiVision CE-3000 *contrast* control was modified, to permit a no-enhancement condition at the lower end of the range (i.e. $K = 1$ at contrast = 27 arbitrary units out of 99). It also enabled a reduction of contrast ($K < 1.0$), resulting in low-pass filtering of the images for contrast settings lower than 27. A setting in this range reduces the contrast of the image details in the spatial frequency range selected by the *Detail* setting. The settings between 27 and 99 defined linearly amplification that increased over a range of 1.0–3.0 times. A setting in this range increases the contrast of spatial details in the range selected by the *Detail* setting. The right/left direction of the graphics tablet was used to control the contrast parameter, with higher contrast on the right, no enhancement at level 27, and contrast reduction (or low-pass filtering) to the left (see *Figure 2*).

When the contrast is set to $C = 27$ the CE-3000 can then be switched manually to the bypass mode. When this is done normally-sighted observers note no change in the image. This was done to verify that the system operated correctly. The bypass video can also be compared to a video that physically bypasses the CE-3000 and these images are also indistinguishable from the original images.

The *background* control (having values ranging from 0 to 99) sets the fraction (linearly between 0 and 1.0) of the low-pass filtered mean luminance image that is added back to the high-pass filtered image. *Background* is indirectly related to the L parameter in Peli and Peli (1984). With background set to 99, none of the original low-pass filtered mean luminance image (except for a uniform gray background) is added to the high-pass



Figure 1. An example of (a) the original, unenhanced image and (b) an image whose detail and contrast settings have been enhanced in accordance with an individual subject's selection. In the experiment, images were presented in colour on a 27" television. Note the moderate level of enhancement selected and changes in local luminance (particularly on the trouser leg) that permit greater enhancement of high frequency details (such as the folds in the material).



filtered image, and with background set to 0 all of the low-pass filtered mean luminance image is added to the high-pass filtered image. Since the position on the graphics tablet provided only for the control of two parameters, the background parameter was implemented in a fixed linear proportion to the contrast parameter set by the graphics tablet. A background value of 0 (i.e. no change in background) was used for the contrast ≤ 27 , and a background of 0.5 was assigned for the maximal contrast parameter (i.e. contrast = 99 arbitrary units, representing enhancement). Background values greater than 0.5 were not used because it was necessary to reduce the

Figure 2. Enhancement parameters chosen by each subject. The position of each cross corresponds to mean mouse position on the graphics tablet and the error bars represent S.E.M. The vertical line at contrast = 27 (arbitrary units) represents no enhancement (original image). Area to the left of that line represents image degradation by low-pass filtering. Area to the right represents enhancement. Area to the right of the shaded diamond represents over-enhancement. All subjects selected values corresponding to enhancement. (a) The combined group of 46 patients, 21 of whom averaged across five images, two criteria, and two repetitions, and 25 of whom averaged across four images and two repetitions. (b) The same data but only for the 20 subjects who completed the continuous evaluation of perceived quality part of the experiment. The filled symbols to the right of the vertical line represent enhancement settings used in the second part of the experiment for one subject. Diamond, individually-selected enhancement settings; Triangles, settings used if in the B_x group; Squares, settings used if in the B₊ group; Open circles, settings that resulted in two degraded images; Filled circle, setting for the unenhanced original image resulted from any setting with contrast = 27.

amount of low-pass filtered details to provide sufficient range (headroom) for the presentation of the amplified high frequency content of the image (Peli, 1992).

Subjects

All 56 subjects had central vision loss, mostly due to age-related macular degeneration (AMD). All subjects were presumed to have central field loss (CFL) based on diagnosis or retinal images and, in all but six, CFL was confirmed by a visual field test. These six subjects were included in the analyses. Subjects signed an informed consent form approved by the Institute Review Board (ethical committee).

The subjects comprised two main groups, summarized in *Table 1*. Group A participated in a preliminary study to refine the procedure for the selection of preferred enhancement parameter settings. Two subjects from this group were unable to manipulate the cursor on the graphics tablet due to arthritis, thus reducing n to 21. Eight subjects from group A were also included as part of group B. Following this preliminary study, group B_p participated in a shorter procedure for the selection of enhancement parameters. One subject in this group could not manipulate the tablet, and three additional subjects who did not have time to finish watching the movie segments were excluded, thus reducing the number who selected preferred enhancement levels to 25. This group went on to complete the performance portion and the continuous evaluation of perceived image procedure as shown in *Table 1*.

Individual selection of enhancement parameters

For the individual selection of the enhancement parameters, the subjects in group A were presented blocks of five

Table 1. The characteristics of the groups of subjects that participated in the various test procedures

Group	Subgroup	Video	n	Age range	Acuity range	Documented CFL
A			21	22–88	6/30–6/180 (20/100–20/600)	18
B	B _p	Mystery	25	43–90	6/30–6/240 (20/100–20/800)	22
	B _x	Maya	10	65–89	6/48–6/120 (20/160–20/400)	9
	B ₊	Maya	10	48–92	6/30–6/150 (20/100–20/500)	10

Group A only piloted the selection of preferred enhancement levels. Group B participated in the performance assessment (group B_p) and the continuous evaluation of perceived quality (groups B_x and B₊).

different static video images digitally frozen using a time base corrector (FA-300, FOR-A Company, Tokyo, Japan). These images were described as: 'On the phone', 'The cook', 'Garden', 'Choir' and 'Tavern at night'. The images were chosen to represent a variety of scenes encountered in movies and in relation to the following motion video presentation. They included: different scenery, different characters, and were selected to ensure that seeing these static images would not influence performance on recognizing details from the motion segments shown later. The five images were presented in four blocks. The order of the five images within each block was randomized. The subjects selected their preferred adaptive enhancement parameters by choosing a mouse position on a graphics tablet. The subjects controlled the 'Detail' parameter, representing spatial frequency, along the up/down dimension of the pad and a combination of the 'Contrast' and 'Background' parameters on the right/left dimension. The subject was asked to move the mouse to explore the whole surface of the graphics tablet while noting the changes in the image on the screen. The subject was then asked to refine a search for the point that provided the optimal image. The subject was encouraged to complete 2–3 iterations of the process before making a decision. The active area of the graphics tablet was randomly shifted up, down, right or left on each trial in order to prevent subjects from choosing a default mechanical position and not properly exploring the options. The subjects were told that the mapping of the enhancement parameters to the graphics tablet was varied from trial to trial so that the same position on the graphics tablet represented a different effect on the image at every trial. In different blocks the subject was asked to select the optimal image based on one of two preference criteria: (1) 'best overall' image (two blocks), and (2) 'most details' visible in the image (two blocks). The selection criteria alternated from 'best overall' to 'most details' for alternating blocks. Thus the total image count for the subjects in group A (Table 1) was 20: five images × two preference criteria × two repetitions of the static images.

Since the different preference criteria resulted in no significant difference between the settings selected by the subjects (see *Results*), later experiments eliminated the blocks with the 'most details' instructions. To enable the entire set of procedures to be completed in a

manageable amount of time, we also reduced the number of images to four by excluding the night scene, for which responses varied the most. Thus for the subjects in group B the total image count was reduced to eight: four images × two repetitions.

Effect of enhancement on performance

The individual average selected enhancement setting for subjects in group B_p (shown in Figure 2a) was determined and then set on the DigiVision CE-3000 for use during the presentation of movie video segments. The video program 'Poirot: The Theft of the Royal Ruby', an episode of the Public Broadcast Service show *Mystery!* was used. Recognition was tested from the video program by measuring the number of visual details that could be correctly identified in response to the questions after observing either the original (not enhanced) or the enhanced video segments. Multiple-choice questions were posed after each short segment. Questions addressed visual details, e.g. 'The woman has... (1) gray hair; (2) black hair', that were described by the original AD prepared for broadcasting by the Public Television station WGBH in Boston for this TV program. The normal program audio was played without the AD. The performance measure used here was the one developed by Peli *et al.* (1996) to assess the effectiveness of AD. AD provides verbal descriptions of the visual elements of TV programs through a third audio channel without interfering with the programs' standard audio portion (Cronin and King, 1990). Descriptions of visual details concerning aspects such as clothing and colours are inserted during pauses in the dialogue. AD is available on DVD media, on videocassette tapes, on public broadcast television in the United States using Descriptive Video Services (DVS), and in the United Kingdom by using the Freeview set-top box called the i-Player AD, endorsed by the Royal National Institute of the Blind (RNIB) and the BBC.

Peli *et al.* (1996) developed questionnaires that asked about details that were described in the AD of three public broadcast programs. The effect of AD was evaluated by administering these questionnaires to partially-sighted subjects who watched the programs with or without the AD. Here, one of these question-

naires was used to assess the effect of image enhancement on performance when subjects viewed enhanced video segments (without AD).

The questionnaire for the 'Mystery!' video consisted of 59 questions dealing with a 10 min segment of the episode. For the performance evaluation, every subject viewed the same video. The video was paused 17 times, at proper break points, to administer the AD-based questions. The initial condition (either enhanced or unenhanced) was counterbalanced across subjects. After the 30th question, the condition was switched. Thus, half the subjects viewed the first part of the segment in the enhanced mode and half the subjects viewed it in the unenhanced mode. The parameters for the enhanced condition were the individually-selected enhancement settings determined in the first part of the experiment. The CE-3000 device was put into 'bypass' mode to present the unenhanced segments.

Following the performance evaluation procedure, a brief second questionnaire was administered regarding the subject's *impression of overall quality* of various aspects of the video segments, as described below.

Impression of overall image quality

To evaluate *impression of overall quality*, following the first condition in the performance study (enhanced or unenhanced), each subject was asked to mark his response to seven questions, comparing the segments he just saw to normal TV viewing. The responses were indicated by moving a marker across a continuously numbered scale, which was labelled by the words 'poor' and 'excellent' at the ends of the scale in large print [the orientation of which (right-left) was counterbalanced across subjects]. The experimenter recorded the subject's responses from the scale (range of 0-50). Comparisons were made of the seven measures: colour, visibility of details, ability to recognize faces, ability to discern facial expressions, ability to follow story, sound quality and overall impression. Following the presentations of the second condition in the performance study, the comparison questions were repeated. In this case, for each question the experimenter positioned the marker to the previous setting selected by the subject for this question and the subject was asked to indicate his response in comparison to his previous selection for the first condition. At the end of the session the subject was asked which of the two presentations was preferred, which appeared processed, and which appeared most like normal TV.

Continuous evaluation of perceived image quality

The procedure to evaluate the *perceived image quality* of motion video was derived from the method Hamberg

and de Ridder (1995) used to evaluate perception of dynamic changes applied to static imagery. The video program 'Lost Kingdom of the Maya', a *National Geographic Special* was used for this part of the experiment.

This method enabled continuous measurement of the perceived image quality as the display parameters (details and contrast/background) were changed. Subjects indicated perceived quality related to the adjectives: excellent, good, sufficient, poor and bad by moving the mouse on a scale printed in print large enough to be read easily by the subjects. An audio beep every 10 s indicated to the subject a change in parameters and mouse position selected in response to the new view was recorded (1 per second). The motion video segments played continuously (with no sound) while the enhancement parameters were abruptly switched between the following sets of eight possible values:

1. The individually-selected set of parameters: Detail, D , and Contrast, C , chosen as described above, and indicated in *Figure 2b*. In particular for one subject, this is illustrated in the figure by a shaded diamond at this data point.
2. The original image (D , $C = 27$), indicated for that same subject with a dark circle on the vertical line. Note that the value of parameter D has no effect in this case.
3. Two settings that degrade the image, rather than enhance it, by low-pass filtering at two different levels of (D , $C < 27$) (corresponding to contrast gains of 75 and 37%, respectively). The positions of these settings are represented for the sample subject by the two open circles to the left of the vertical line in *Figure 2b*. Note that for these settings, the detail parameter, D , was the individually-selected detail parameter.
4. Intermediate and over-enhanced values for both parameters: (D , $C/2$); (D , $C \times 2$); ($D/2$, C); ($D \times 2$, C). The actual value for the setting designated $C/2$ is $(C - 27)/2 + 27$ and the actual value for the setting designated $C \times 2$ is $(C - 27) \times 2 + 27$. The positions of these settings are indicated by squares in *Figure 2b*. We refer to these arbitrary enhancements as the plus (+) arbitrary enhancements because the spatial arrangements of these enhancements in *Figure 2b* form an upright cross or a plus sign shape. The 10 subjects using these settings are referred to as the B_+ group. For the other 10 subjects (the B_\times group), a different set of intermediate and over-enhanced values: ($D/2$, $C/2$); ($D \times 2$, $C \times 2$); ($D/2$, $C \times 2$); ($D \times 2$, $C/2$) were used. The positions of these settings are indicated by triangles in *Figure 2b*. We referred to these settings as the crossed (\times) enhancements.

For both groups of subjects, each of the eight settings was randomly presented within 10 blocks, for a total of 80 presentations.

Data analysis

Analysis of performance data

The percent correct responses on the AD-based questionnaire were tallied for each subject for the enhanced segments and the unenhanced segments. Two of the 59 questions were excluded from this analysis because more than 90% of the subjects in the prior study (Peli *et al.*, 1996) were able to answer them correctly from just listening to the audio portion of the video segment without viewing it at all. Another question was eliminated because only 10% of the normally-sighted group in that study was able to answer it correctly. The overall means of the number of questions answered correctly for the enhanced segments and unenhanced segments were compared using paired sample statistics.

The data from the *impression of overall quality* test consisted of the marker settings of the enhanced and unenhanced conditions. These were converted to a score of -1, 0 or +1 depending on whether the subject preferred the original, liked both equally, or preferred the enhanced segment. The scores for the seven questions were then summed to produce an overall preference score that ranged from -7 to +7. Thus a score of -7 would mean that the subject always preferred the original and a score of +7 would mean that the subject always preferred the enhanced. Linear regressions were performed to determine if the overall preference scores were correlated with performance and settings of the enhancement parameters.

Analysis of continuous perceived quality data

As the subject viewed the video, the quality score as indicated by the mouse position was recorded at 1-second intervals for 10 s. The first 3 s of data were considered the 'transition period' during which the subject moved the mouse from one value to another in response to a change in the enhancement parameters. The last seven data records from each sequence were averaged to obtain the scores for that detail/contrast combination. Each condition was repeated 10 times and the scores for each of these repetitions were converted to the probabilities used for the ROC analysis (see below).

ROC analysis

Data were analysed using a signal detection approach (Macmillan and Creelman, 1991). The ROCKIT program (Metz *et al.*, 1984) was used to determine the fitted area under the receiver operating characteristic (ROC) curve (A_z) (Hanley and McNeil, 1982). Paired comparisons were made between responses to the original video

segments and a set of processed video segments. As there were seven sets of processed segments for each subject, seven ROC curves were determined that represented the difference in perceived image quality between the original and the enhancement done with that set of enhancement parameters.

In ROC analysis, a detector's (i.e. subject's) responses to 'noise' presentations and to 'noise-plus-signal' presentations are compared. In the current study, for the purpose of the ROC analysis, the original images were treated as the 'noise presentations' and the processed images were treated as the 'noise-plus-signal presentations'. Subjects were asked to report perceived image quality, so they could be considered image-quality detectors. As can be seen in *Figure 3*, our raw data consisted of multiple distributions along the perceived image quality dimension (for clarity *Figure 3* only shows data for four of the seven types of processing presented to the subject). The program controlling the graphics tablet produced a score in the range 0–5000. To produce *Figure 3*, the 7-second averages were binned into 10 equal bins. However, this binning was not used for the ROC analysis. To produce the data points used in the ROC analysis (*Figure 4*), two ratios were computed for a given score: (1) the ratio of the number of responses to the original images that had a greater score than that given score, to the total number of original image presentations (O-proportion) and (2) the ratio of the number of responses to the processed images that had a greater score than that given score, to the total number of processed image presentations (P-proportion). The pairs of ratios were used as the (x , y) coordinates of the ROC data points. These were then fit to an analytic cumulative normal function with two parameters (Metz *et al.*, 1984). As our ROC analysis was of perceived video quality – not of enhancement detection, as is usually done – the traditional labels of the axes of the ROC figure (e.g. true-positive rate, or 'hit' rate) do not directly apply to our situation.

The area under the analytic fit curve (A_z) was taken to be a measure of perceived quality and used in subsequent statistical tests. For example, when the perceived image quality of the processed videos ($D = 46$, $C = 62$) was better than the original (46, 27) video set in *Figure 3* the distribution is further to the left, and the resulting A_z was greater than 0.5 (*Figure 4*). For the degraded video set, the perceived video quality was worse and the distribution was to the right of that of the original, resulting in $A_z < 0.5$.

The ROCKIT program provides 95% confidence limits for each A_z (Hanley and McNeil, 1982; Metz *et al.*, 1984), and where appropriate these are reported. The confidence intervals were used to determine the significance of the responses of individual subjects to a particular type of image processing (i.e. A_z was

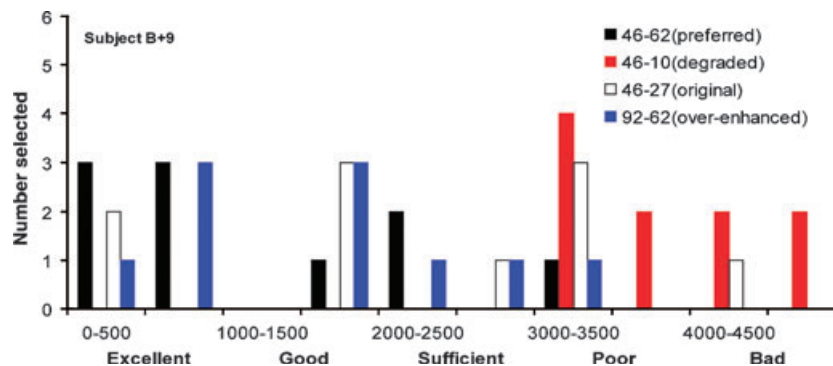


Figure 3. Frequency distribution of the continuous perceived image quality scores indicated by a subject using the graphic tablet in response to four of the test video segments: the original unprocessed video (setting 46-27 on the Digivision), the individually-preferred enhancement (setting 46-62 for this patient), a degraded video (46-10) and an over-enhanced video (92-62). The subject responses were binned into 10 bins of 500 arbitrary graphic tablet units. This subject clearly preferred his individually-selected enhancement and thus has one distribution (46-62) that is shifted to the left of the others. The over-enhanced distribution (92-62) is also shifted to the left, but not as much as the preferred enhancement. The distribution for the degraded set is highly shifted to the right. For simplicity, only four of the eight distributions (corresponding to the seven enhancement conditions and the original image tested for each subject) obtained are shown. Distributions from a subject such as shown in *Figure 4b* would not be clearly separated. The scores shown here were used to construct three corresponding ROC curves shown in *Figure 4a*.

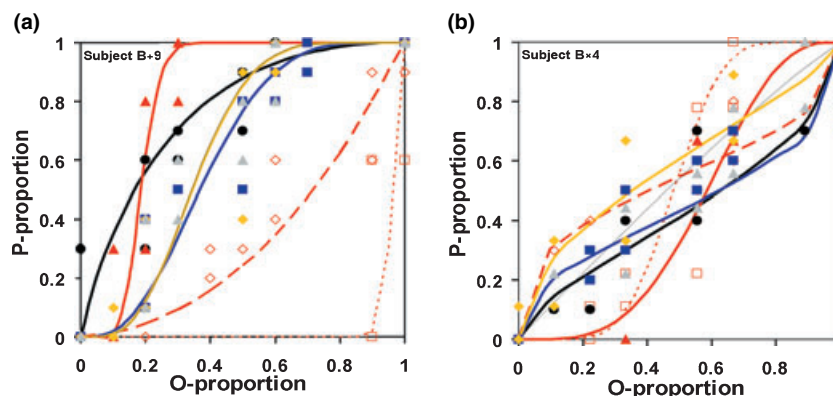


Figure 4. ROC data and fitted curves for differences in perceived image quality between the original and processed video segments. (a) Data for one subject (9+ from group B₊) who preferred all of the enhancement settings, but the perceived image quality was only significantly different from the original for the individually-selected enhancement (circles) and one of the arbitrary enhancements (triangles). This subject rejected the degraded segments (open squares and diamonds and dashed lines) as compared to the original segments (but only the most degraded was significantly rejected). The colour-coding of the data and ROC curves corresponds to that used in *Fig. 3*. (b) This example shows a subject who did not perceive much difference in image quality between the processed video segments, including both the degraded and enhanced videos and the original video.

considered significantly different from 0.5 when the 95% confidence interval did not include 0.5).

In all analyses, significance of $p < 0.05$ was considered a statistically significant difference. Unless otherwise stated, all variances are reported as standard error of the mean (S.E.M.). Correlations and ANOVAs were performed to see if there were effects of enhancement parameters or of subject group.

Results

Individual selection of enhancement

For group A, a univariate five (images) \times two (criteria) \times two (repetitions) ANOVA on the *contrast* setting

showed no effect of criteria ($F_1 = 1.47, p = 0.23$) or of repetition ($F_1 = 1.03, p = 0.31$) but did reveal an effect of image ($F_4 = 6.4, p < 0.005$). Similarly, for the *detail* setting, there was no effect of criteria ($F_1 = 0.30, p = 0.58$) or repetition ($F_1 = 0.15, p = 0.70$) but an effect of image ($F_4 = 2.89, p = 0.02$). There were no significant interactions between the three factors. The strong effect of image indicates that there are different settings for different images. This might be anticipated in consideration of the different spatial content that is relevant and interesting in different images. Since there was no main effect of repetition or criteria, the individual selection of parameters for group B was reduced to two repeated blocks, four images, using only the ‘best overall’ criterion.

Figure 2a presents the mean and S.E.M. of the enhancement parameters selected by each of the 46 subjects from both groups who were able to manipulate the cursor of the graphics tablet. Subjects never selected unenhanced or low-pass filtered images. These findings indicate that we can measure preference for enhancement with static images, and that subjects always preferred the enhanced to the unenhanced images. The data show a significant correlation between the two parameter settings such that as the Detail setting (spatial frequency of the enhancement) decreased, the Contrast level selected increased ($r = -0.65$, $p = 0.0005$). Figure 2b shows the same data, but only for the 20 subjects (groups B_x and B₊) who participated in the continuous evaluation of perceived quality part of the study. For this smaller group of subjects, the correlation between the Detail and Contrast was not significant ($r = -0.25$, $p = 0.30$), though the trend is similar. For the 46 subjects who selected a preferred enhancement setting, neither Contrast ($r = 0.14$, $p = 0.35$) nor Detail ($r = 0.09$, $p = 0.55$) settings were significantly correlated with subject's visual acuity (in logMAR); neither were the Contrast ($r = 0.02$, $p = 0.88$) or Detail ($r = 0.04$, $p = 0.78$) settings correlated with subject age.

Effect of enhancement on performance

Twenty-five subjects (group B_p) completed the AD-based questionnaire performance evaluation. The subjects were able to answer 66% of the questions correctly when the video was presented enhanced and 71% when it was presented with no enhancement. This difference was not statistically significant (paired sample *t*-test, $t_{24} = 2.04$, $p = 0.053$). Note that in agreement with our previous results (Peli *et al.*, 1996), the subjects could answer over 70% of the questions correctly without enhancement and without hearing the DVS description, leaving very little room for potential improvement (ceiling effect).

In response to the single question 'which presentation do you prefer' (that was asked after the performance evaluation), 14 subjects preferred the enhancement overall while nine preferred the unenhanced video (two subjects did not complete the questions because of time constraints). This difference was not significant ($\chi^2 = 1.09$, DF = 1, $p = 0.30$); Table 2 shows the performance scores for those who preferred and those who did not prefer the enhanced videos. There was no significant difference in the scores on the enhanced videos between the two groups (*t*-test for equality of means, $p = 0.30$). However, for those who preferred the unenhanced videos, there was a significant difference in their scores on the enhanced vs unenhanced videos ($p = 0.037$), which was consistent with their preference.

For each of the seven questions on preference, a score of -1, 0 or 1 was given to indicate whether the subjects

Table 2. Percentage of questions correctly answered for group B_p (mean \pm S.D.)

Preferred video	Score on enhanced video (% correct)	Score on unenhanced video (% correct)
Enhanced ($n = 14$)	70 \pm 12	69 \pm 11
Unenhanced ($n = 9$)	64 \pm 12	78 \pm 6

Subjects who preferred the enhanced video did not perform better on the enhanced segments, but subjects who preferred the unenhanced images performed better on these segments.

preferred the unenhanced, liked both equally, or preferred the enhanced. The overall mean of these scores was 0.15 ± 0.08 (S.E.M.), which indicated a slight preference for the enhanced images, that only approached statistical significance (one sample *t*-test, $t_{111} = 1.89$, $p = 0.062$). The mean preference scores for each subject were not significantly different from their single answers on preference (paired samples *t*-test, $t_{15} = -0.72$, $p = 0.48$).

Continuous perceived quality with motion video

Two groups each with 10 subjects (Table 1, groups B₊ and B_x) participated in the continuous evaluation of perceived image quality testing. For both groups each subject was presented with her/his individually-selected enhancement, with the original unenhanced segments, and with two levels of degraded images. Each subject was also presented with sets processed with four additional arbitrary enhancement levels, two of which were over-enhanced. For one of these groups the plus (+) configuration of the arbitrary enhancement parameters was used and for the other group the crossed (x) configuration was used.

The ROC analyses with the fitted curves for two representative subjects are presented in Figure 4. The results of the ROC analysis for one subject (from the B₊ group for all conditions are shown in Figure 4a and summarized in Figure 5. For this subject the area under the ROC, A_z , was significantly higher than 0.5 for the individually selected enhancement and for only one of the arbitrary enhancements. The A_z was significantly lower than 0.5 for the highly degraded setting, but not for the moderately degraded setting. For all other conditions, the 95% confidence intervals for the A_z included the 0.5 levels and therefore indicated that the perceived image quality for these conditions was not significantly different from the perceived quality of the original unenhanced video. The results of the ROC analysis for a different subject (from the B_x group) for all conditions are presented in Figure 4b. For this subject, the area under the ROC was not statistically significantly different than 0.5 for any of the conditions (the 95% confidence interval included the 0.5 for all

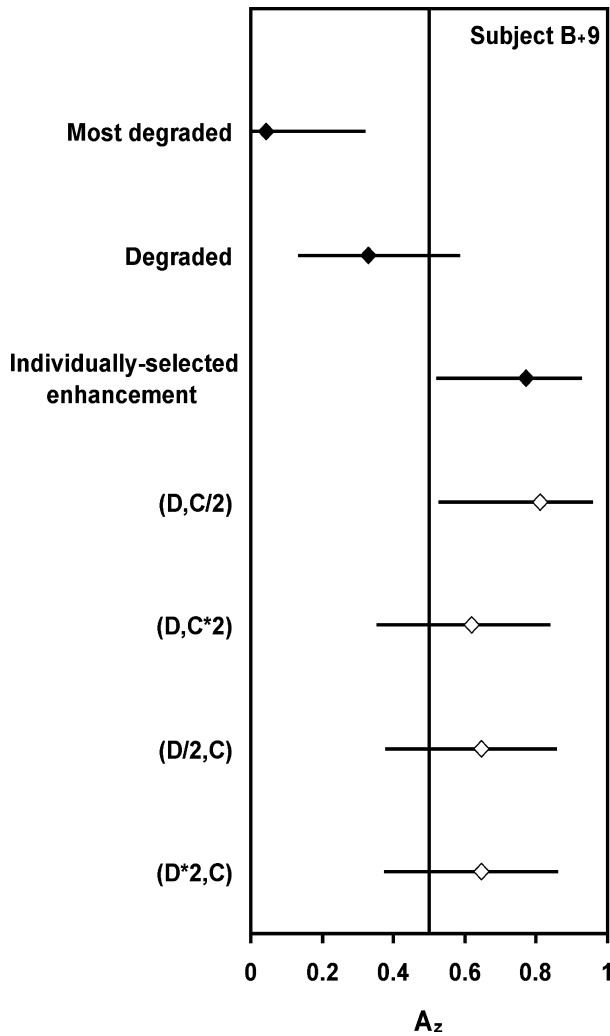


Figure 5. A_z (area under ROC curve) values for subject 9+, whose ROC curves are shown in Figure 4a. A_z values were significantly higher than 0.5 for the individually selected enhancement and one of the arbitrary enhancements indicating that perceived image quality for these two enhancements was significantly different from the original. Error bars are 95% confidence levels as determined from the ROCKIT program.

conditions). These results indicate that this subject (acuity 20/160) did not perceive a significant difference in quality between the original video segments and the processed video segments, the degraded videos, or the enhanced videos.

Since both groups of subjects [the plus (+) and crossed (x)] were presented with the individually-selected enhancement and with the same levels of degraded images, these results were averaged for the two groups together, while the arbitrary conditions were averaged separately for each group. The results (Figure 6) demonstrate that the subjects preferred the enhanced videos to the unenhanced videos ($t_{107} = 6.92, p < 0.0005$) and preferred the unenhanced videos to the low-pass filtered videos ($t_{41} = -4.06, p < 0.0005$). The average response

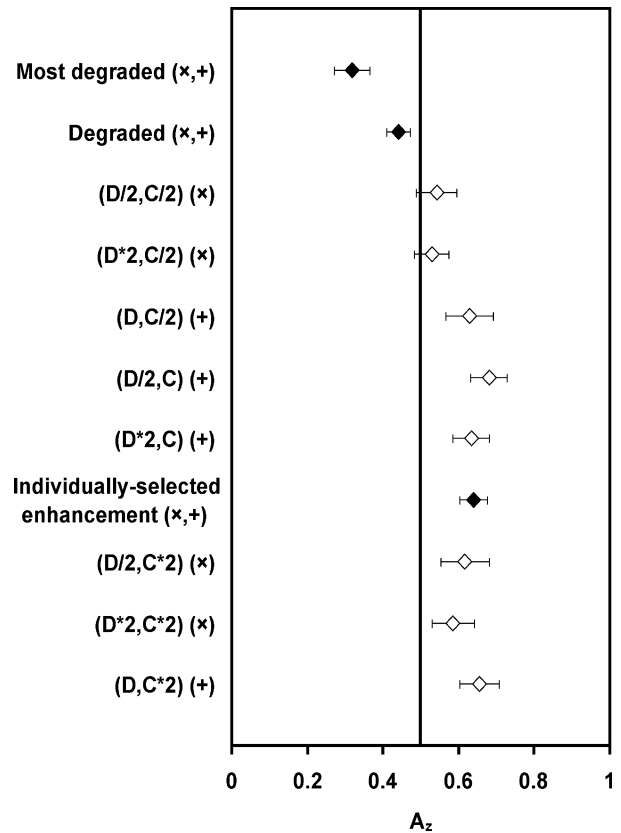


Figure 6. Average A_z for all subjects. The three filled points show the two degraded conditions and the individually selected enhancement condition, which were common across all 20 subjects. The degraded conditions have average A_z less than 0.5 whereas the enhanced condition has an average A_z greater than 0.5. The other points show all the other enhanced conditions for the (B₊) and the (B_x) groups of 10 patients each. The (x) and (+) symbols refer to the set of conditions presented as described in the text. For all except two of these enhancement conditions, the A_z is significantly greater than 0.5 as indicated by lower bound of the error bars (S.E.M.).

for the original is equivalent to the 'Sufficient' setting and the average response of the individually-selected enhancement is equivalent to the 'Good' setting. Additionally, individually-selected enhancement resulted in statistically significant improvement in perceived quality ($A_z = 0.64, \pm 0.17$) over the unenhanced images (0.5) (one sample t -test, $DF = 21, p = 0.001$). No differences in perceived quality were found between the individually-selected set of parameters and the corresponding arbitrary enhancement values in either group.

For the combined group of 20 subjects, we found no significant correlation of the summed preference values (-7 to +7) with the level of Contrast selected (Pearson's $r = 0.30, p = 0.13$) or with the level of Detail selected (Pearson's $r = -0.20, p = 0.23$). However, there was a significant correlation of the preferred Contrast setting with the A_z (Pearson's $r = 0.57, p = 0.004$), but not the Detail setting (Pearson's $r = 0.024, p = 0.46$).

Discussion

We have found a statistically significant effect of enhancement in a continuous measure of preference for motion video. These results further illustrate that the adaptive enhancement (Peli and Peli, 1984) as provided by the DigiVision CE-3000, adds significantly to perceived image quality by visually impaired patients. The fact that a small group of low vision subjects ($n = 20$) is sufficient to demonstrate a statistically significant effect is an indication that the effect is robust. This finding is also supported by the fact that of the 25 subjects in group B_p, 17 correctly identified the enhanced video as processed, and only six erroneously identified the unenhanced as processed (two did not complete the questions due to time constraints). The six misidentifications could be attributed to one of three possible explanations: (1) It could have been due to a misunderstanding of the instructions/questions on this item. The subjects in this study were mostly old and previously unfamiliar with the concept of processed images. (2) It could be that the difference between the two presentations was so small that it was not really noticeable. This is an unlikely explanation as all subjects clearly selected an enhanced presentation as their preferred presentation. If the difference was not noticeable some might have selected no enhancement or even degraded presentation, as did a few subjects in another study (Tang *et al.*, 2004). (3) It is possible that the response of these subjects represents an adaptation to the enhanced images, leading them to judge the enhanced images as looking natural and thus perceiving the original video to be blurred (i.e. processed). Webster *et al.* (2002) reported such an effect. They concluded that visual responses are continuously calibrated to compensate for variation in sensitivity with spatial scale. If this is the case, then any image enhancement might lead to adaptation, reducing the perceived benefit after short-term use. Such adaptation, however, might be counteracted by leaving a part of the image (a frame or a margin) unprocessed, permitting the initial calibration to be maintained.

As a group, the subjects indicated only a modest change in perceived quality of the images under all manipulations. While the quality of the original image was rated on average as sufficient, the most degraded image, which was quite severely degraded, was rated less than one step below that (better than poor, on average). At the same time we have seen a very consistent selection of enhancement parameters during the selection and preference parts of the experiments. There may be a few explanations for these seemingly contradictory results. One possible reason is that the selection of enhancement parameters was done with static images while performance testing used live video. It may also be related to the main effect of image that we found in the

selection procedure. It is possible that the enhancement selected for these static images was less than optimal for the motion videos, which purposely included scenes different from those used in the selection process. The interaction we found between image and enhancement parameters might mean that to be most effective the enhancement would have to be continuously tuned (in real time) by the users of such a system (or by some automatic process in response to the changing characteristics of the images). At this time, however, we do not know what these characteristics might be. We are investigating the changes in spatial frequency content of video sequences, as one possible variable.

The variability in results found across our various studies with the adaptive enhancement may be accounted for by the changes in the test methodology and possibly by variability in the group's population. Visually impaired patients are not a uniform group, and the sampling effect may be much larger in studying these patients than those commonly found in psychophysical studies with normally-sighted observers.

The question of individual vs generic enhancement has not been answered definitively by this study. The variations between various levels of enhancement used in the study did not lead to statistically different effects (Figure 6). The individual selection did lead to a larger effect than all other conditions and the lack of significant difference may be due to the small number of subjects or the limited level of variation used (and possible with this device and in video in general). It is clear, however, that subjects with low vision can select a level of preferred enhancement that is consistent (small S.D. on repeated selection and across images) and follows a regular pattern for the group as a whole. We can also conclude that the enhancement selected using static images is preferred for the viewing of motion video. Subjects who preferred stronger enhancement (larger settings of the Contrast parameter) also perceived the enhancement to provide a larger benefit in image quality (larger A_2).

Using the AD-based questionnaire method of testing, we were not able to demonstrate an improved performance in recognizing details from video when using the enhanced images. A prior study that used the same methodology and questionnaire to assess the effect of AD itself (Peli *et al.*, 1996) found that there was considerable individual variability and that, as we found here, many low vision subjects were able to correctly answer many of the questions even without enhancement and without hearing the AD description. Thus, this method could detect improvement in performance only if the effect of enhancement was relatively large. Using AD-based questions might not be a sufficiently sensitive measure to the level of performance on the complex task of watching and perceiving a video program. The AD description, having been designed

for blind as well as low vision audiences, therefore underestimates the ability of low vision subjects to perceive many details of the programs, especially from the short distance they use for viewing. The complexity of the task itself, the use of program auditory information, context, and individual background and interests, make the testing of perception of details from a movie a particularly difficult task. Thus, while the potential of image enhancement for the visually impaired appears promising, the best methodologies to assess the effects of enhancement on perceived image quality and, even more so, on recognition performance is still to be developed. Such test methodologies are needed to enable improvement in design and selection of parameters in any enhancement approach.

Acknowledgements

Supported by NIH grants EY05957, and EY10786. The author thanks Angela Labianca for her help on many aspects of the study. Charles Simmons programmed the graphics tablet control of the enhancement device. Bob Goldstein and Laurel Bobrow analysed the ROC data and prepared the figures. Rick Hier from DigiVision provided the specially modified system that low-pass filters the images as well as enhances them. Barry Cronin, Gerry Field and Laurie Everett of WGBH, Boston provided the program videos, software and invaluable technical help.

References

- Bryant, R. C., Lee, C. M., Burstein, R. A. and Seibel, E. J. (2004a) Engineering a low-cost wearable low vision aid based on retinal lighting scanning. *SID Dig.* **35**, 1540–1543.
- Bryant, R. C., Seibel, E. J., Lee, C. M. and Schroder, K. E. (2004b) Low-cost wearable low-vision aid using a handmade retinal light-scanning microdisplay. *J. SID* **12**, 397–404.
- Cronin, B. J. and King, S. R. (1990) The development of the descriptive video service. *J. Vis. Impair. Blind.* **84**, 503–506.
- Everingham, M. R., Thomas, B. T. and Troscianko, T. (2003) Wearable mobility aid for low vision using scene classification in a Markov random field model framework. *Int. J. Hum. Comput. Interact.* **15**, 231–244.
- Fine, E., Peli, E. and Brady, N. (1997) Video enhancement improves performance of persons with moderate visual loss. In: *Proceedings of the International Conference on Low Vision, "Vision'96"*, Organización Nacional de Ciegos Españoles, pp. 85–92.
- Fine, E. M. and Peli, E. (1995) Enhancement of text for the visually impaired. *J. Opt. Soc. Am. A* **12**, 1439–1447.
- Hamberg, R. and de Ridder, H. (1995) Continuous assessment of perceptual image quality. *J. Opt. Soc. Am. A* **12**, 2573–2577.
- Hanley, J. A. and McNeil, B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36.
- Hier, R. G., Schmidt, G. W., Miller, R. S. and DeForest, S. E. (1993) Real-time locally adaptive contrast enhancement: a practical key to overcoming display and human-visual-system limitations. *SID 93 Dig.* **24**, 491–494.
- Isenberg, L., Luebker, A. and Legge, G. E. (1989) Image enhancement for normal and low vision. *Invest. Ophthalm. Vis. Sci.* **30**(Suppl. 4), 396.
- Kim, J., Vora, A. and Peli, E. (2004) MPEG-based image enhancement for the visually impaired. *Soc. Photo-Opt. Eng.* **43**, 1318–1328.
- Lawton, T. B. (1992) Image enhancement filters significantly improve reading performance for low vision observers. *Ophthalmic Physiol. Opt.* **12**, 193–200.
- Macmillan, N. A. and Creelman, C. D. (1991) *Detection Theory: A User's Guide*. Cambridge University Press, Cambridge.
- Metz, C. E., Wang, P. and Kronman, H. B. (1984) A new approach for testing the significance of differences between ROC curves measured from correlated data. In: *Proceedings of the Eighth Conference on Information Processing in Medical Imaging* (ed. F. Deconinck). Martinus Nijhoff, pp. 431–445.
- Myers, L. R. Jr, Rogers, S. K., Kabrisky, M. and Burns, T. J. (1995) Image perception and enhancement for the visually impaired. *Eng. Med. Biol. Mag. IEEE* **14**, 594–602.
- Omoruyi, E. J., Leat, S. J., Kennedy, A. and Jernigan, M. E. (2001) Image enhancement filters for the visually impaired: a comparison of generic and customized filters. *Vis. Sci. Appl. OSA Tech. Dig. Ser.* **10-3**, 64–67.
- Peli, E. (1992) Limitations of image enhancement for the visually impaired. *Optom. Vis. Sci.* **69**, 15–24.
- Peli, E., Arend, L. E. and Timberlake, G. T. (1986) Computerized image enhancement for low vision: new technology, new possibilities. *J. Vis. Impair. Blind.* **80**, 849–854.
- Peli, E., Fine, E. M. and Labianca, A. T. (1996) Evaluating visual information provided by audio description. *J. Vis. Impair. Blind.* **90**, 378–385.
- Peli, E., Fine, E. M. and Pisano, K. (1994a) Video enhancement of text and movies for the visually impaired. In: *Low Vision: Research and New Developments in Rehabilitation*, Vol. 11. (eds A. C. Kooijman, P. L. Looijestijn, J. A. Welling and G. J. van der Wildt), IOS Press, Amsterdam, pp. 191–198.
- Peli, E., Goldstein, R. B., Young, G. M., Trempe, C. L. and Buzney, S. M. (1991) Image enhancement for the visually impaired: Simulations and experimental results. *Invest. Ophthalmol. Vis. Sci.* **32**, 2337–2350.
- Peli, E., Kim, J., Yitzhaky, Y., Goldstein, R. B. and Woods, R. L. (2004) Wideband enhancement of television images for people with visual impairments. *J. Opt. Soc. Am. A* **21**, 937–950.
- Peli, E., Lee, E., Trempe, C. L. and Buzney, S. (1994b) Image enhancement for the visually impaired: the effects of enhancement on face recognition. *J. Opt. Soc. Am. A* **11**, 1929–1939.
- Peli, E. and Peli, T. (1984) Image enhancement for the visually impaired. *Opt. Eng.* **23**, 47–51.
- Peli, T. and Lim, J. S. (1982) Adaptive filtering for image enhancement. *Opt. Eng.* **21**, 108–112.

- Tang, J., Kim, J. and Peli, E. (2004) Image enhancement in the JPEG domain for people with vision impairment. *IEEE Transact. Biomed. Eng.* **51**, 2013–2023.
- Tu, Z., Chen, X., Yuille, A. L. and Zhu, S. C. (2003) Image parsing: unifying segmentation, detection, and recognition. *Proc. Ninth IEEE Int. Conf. Comput. Vis.* **1**, 18–25.
- Webster, M. A., Georgeson, M. A. and Webster, S. M. (2002) Neural adjustments to image blur. *Nat. Neurosci.* **5**, 839–840.
- Zur, D. and Ullman, S. (2005) *The Low Vision Enterprise. Image Enhancement for the Visually Impaired*. <http://www.wisdom.weizmann.ac.il/~helman/#list1> Accessed on 2 September 2005.