EVALUATING VIDEO ENHANCEMENT FOR VISUALLY IMPAIRED VIEWERS

ELISABETH M. FINE, ELI PELI AND NUALA BRADY

E. M. Fine. Wilmer Eye Institute; Johns Hopkins University, 550 N. Broadway, 6th floor, Baltimore, Maryland 21205
E. Peli and N. Brady. The Schepens Eye Research Institute, Harvard Medical School, 20 Staniford St., Boston, Massachusetts 02114. USA.

INTRODUCTION

For people with visual impairments, the ability to identify visual details is limited due to losses in visual acuity and contrast sensitivity. This becomes particularly problematic when they are engaged in such tasks as reading and face recognition. In 1984, Peli and Peli (see also Peli *et al.*, 1986) proposed contrast enhancement as a means of improving detail perception for visually impaired populations. Peli *et al.* (1991) tested the effects of contrast enhancement on face recognition. Visually impaired viewers were asked to judge whether a face was that of a celebrity. Most (85%) of the patients tested increased their face recognition performance when the images were enhanced.

Until recently, constrast enhancement has been limited to static monochromatic images and offline processing. With the introduction of the DigiVision™ device, which implements a version of the adaptive enhancement algorithm of Peli and Lim (1982; see also Peli & Peli, 1984) on line, contrast enhancement can now be applied to moving images. Television, and other video outlets (e.g. video-taped recordings, computer displays, etc.), are an important source of information and provide access to the culture of a society. While their perception of the visual information is reduced, most visually impaired people enjoy watching TV with their families, and prefer television watching to other activities (Josephson, 1968; Berkowitz *et al.*, 1979).

At this conference in 1993, we (Peli, Fine & Pisano) reported on a preliminary study that was designed to evaluate the efficacy of contrast enhancement for improving perception of motion video among a group of visually impaired viewers. In that study, visually impaired observers used the DigiVision to tun the adaptive enhancement algorithm (Peli & Lim, 1982) to maximize its benefits (defined by each viewer). The algorithm primarily boosts the contrast of the high spatial frequency components of the image. Each subject tuned the enhancement parameters (primarily spatial frequency and level of enhancement) individually using static images taken from the movie. We argued that if the range of spatial frequencies chosen by individual visually impaired patients matched the range at which they had partial loss in sensitivity, and are critical to the perception of detail, the tuning would provide maximal benefit. Allowing observers to make such modifications, we found a significant increase in detail perception with enhancement, as well as an overall preference for the enhanced presentation. However, as noted there, this was primarily a qualitative assessment: viewers were asked to indicate whether the enhanced or natural presentation allowed for easier face recognition, was more natural, etc. In addition, we asked these viewers to answer specific questions regarding some visual details contained in the scenes. However, we had no way to decide which were the important visual details, and we found that the questions we posed addressed visual details that many patients correctly identified without enhancement.

In an attempt to solve the problem of defining which visual details are unavailable to the visually impaired, we used the scripts developed by DVS® to generate questions about visual details in a television program. DVS (and other audio description [AD] services) uses a voice-over technique to describe the details of a given scene during dialogue breaks. An example of such a script is given in Fig. 1. The goal of the description is to convey visual information related to the program that has been deemed important to blind and visually impaired audiences. This determination is made on the basis of focus groups and specific training for developers of AD. While it remains to be determined whether, in fact, these are *important* details, it does provide us with some consensus with this regard, as well as a more formal way of defining the visual information contained in a scene.

Using these scripts we (Peli *et al.*, 1996) developed a series of questions that asked viewers to identify items whose descriptions were contained in the AD. For example, on the basis of the script presented in Fig. 1, we might have asked the question: «Is the man wearing a coat or is he in shirt sleeves?» In our initial study, we asked visually impaired (Snellen acuity worse than 20/100 [6/30]) and normally sighted (acuity 20/40 [6/20] or better) older observers to answer these questions after viewing the segments of the video from which the questions had been developed. When we looked at average performance of the two groups, we found that a group of normally sighted viewers were able to answer 10-15% (depending on the movie) more of the questions correctly than was a visually impaired group of viewers. These data indicate that this method is at least sensitive enough to determine performance differences between normally sighted and visually impaired viewers. Therefore, we used this same technique here to more formally assess the ability of contrast enhancement to increase detail perception from motion video.

In a study using static face images, Peli *et al.*, (1994a) found no difference in performance when each viewer was given the opportunity to tune the enhancement to a preferred setting. This was surprising given that the loss in sensitivity (in terms of visual acuity) was not constant across subjects, nor in the context of the spatial frequency model underlying the enhancement algorithm (Peli & Peli, 1984). However, the range of acuities within the study group was limited, as was the range of preferred settings they chose. The spatial frequency range and the level of enhancement chosen was also very similar to the enhancement pre-selected in the Peli *et al.* (1991) study. This could account for the similarity in performance when the faces were enhanced with pre-selected settings and those of the individual viewers. The use of fixed enhancement could simplify the use of such an enhancement device, and possibly reduce its cost substantially. The use of fixed enhancement also simplifies and shortens the experimental procedure. Because of these issues, we chose to attempt to measure the effects of fixed enhancement on performance and preference for motion videos (movies).

[«... BY A GRANT FROM THE MOBIL CORPORATION.»] On a tombstone, a word appears [NOTES] «Mystery!» A skull winks at us [CLICK], then becomes a keyhole. We see a dead man lying on a library floor, then an elegant ballroom. A woman wears a bat-wing headdress. [THUNDER BEGINS] Outside, a blackcloaked man plays croquet in the rain. [THUMP]

Figure 1. Example of an audio description taken from DVS®. The scene described here is the opening credits for the PBS series *Mystery!*. The text in square brackets represents information from the soundtrack.

We asked a different group of visually impaired observers to watch one of the videos used in the prior study and answer the related questions. They saw half of the video without any processing, while the other half of the video was processed using the DigiVision device. In addition to assessing performance with the contrast enhancement, we also asked our subjects to make qualitative judgments regarding their perception of the video. They were asked to indicate how much difference there was in their ability to, for example, recognize details in the scenes of follow the story (among other variables), when the video was presented under normal viewing conditions and when it was enhanced.

METHODS

SUBJECTS

Nineteen patients from the Schepens Retina Associates participated in this study. Snellen acuity (measured monocularly in the laboratory using a Mentor B-VAT II) in their better eye ranged from 20/100 [6/30] to 20/500 [6/150] with a mean best acuity of 20/214 [6/64] (mdn = 20/200 [6/60]). All had a diagnosis of macular degeneration with central field loss, and some also reported having cataract. They ranged in age from 55 to 90 years (mean = 73; mdn = 72). Participation was voluntary.

APPARATUS

The video taped movie was played through a 3/4 inch high quality video cassette recorder (Sony U-matic SP) and displayed on a 27 in. Sony color television. The image enhancement was controlled via a DigiVision apparatus that implements a version of the adaptive enhancement algorithm of Peli and Lim (1982) in real time. The three variables of the enhancement algorithm, *detail, contrast and background,* are controlled separately by the DigiVision apparatus. *Detail* corresponds to the size of the window over which local luminance is averaged, *contrast* to the amplification of the highpass component, and *background* to the value of the low frequency component. (Specific details regarding the algorithm can be found in Peli & Lim, 1982 and Peli & Peli, 1984; a schematic of the DigiVision implementation is shown as Fig. 2 in Fine & Peli, 1995.) All three variables were fixed at the average preferred settings of low vision observers in our previous study of video enhancement (Peli *et al.*, 1994b). The *detail* setting of 7 corresponds to a 112 x 112 pixel (Gaussian) window (enhancing details at frequencies higher than 4 cycles/image); the *contrast* setting of 7.2 corresponds to an enhancement factor which is 2.4 times the original contrast (similar to that used in the studies with static images); the mean *background* setting of 1.8 corresponds to an 18% increase in luminance relative to the original. (This parameter serves to slightly brighten very dark portions of the image).

The VCR timing was controlled via an Apple Macintosh SE computer using a Hyperecar® stack (de**Scriptor** version 1.2.1; DVS, Boston, MA). In addition to segmenting the program, the software also presented the experimenter with the appropriate questions, and provided space for recording the subjects' responses. These were then transferred automatically to a spreadsheet for further analysis.

PROCEDURE

The observers watched the first 10 min. of «Poirot: the Theft of the Royal Ruby», an episode of the Public Broadcasting Services' *Mystery!* series, that had been aired with AD. We chose this video because it had the greater difference in performance between the normally sighted and visually impaired groups in our previous study (Peli *et al.*, 1996). A total of 59 two-alternative, forced choice questions were developed from the portion of the AD script corresponding to the initial 10 min. of the program. The movie was shown in short sequential segments. After each segment, the video was paused, and subjects were asked questions about the material they had just seen. Each viewer answered either the first or second half of the questions from the natural video and the other half from the enhanced video. The order of condition (natural or enhanced) was counter-balanced across subjects.

After viewing each half of the movie, subjects were asked to rate the quality of the display on seven different dimensions: color, visibility of details, ability to recognize faces, ability to recognize facial expressions, ease of following the story, sound, and overall impression. The ratings were made relative to a fixed standard: subjects judged the display quality in the first half relative to watching television at home, and then judged the display quality in the second half relative to the first half. They indicated their ratings by moving a sliding knob either rightward (to indicate «better than») or leftward (to indicate «worse than»). Leaving the knob in the same position indicated «no difference». After the first half of the movie, the knob was initially positioned in the center of the scale for all seven questions (representing TV at home—the baseline comparison). After viewing the second half, ratings of the different dimensions were made relative to each subjects' rating for the first half of the video. The knob was initially positioned at the point on the scale indicating the subjects' judgment in the first half (the scale was marked with numbers only visible to the experimenter), and they were again asked to move it to the right if the current viewing conditions were better, and to the left if they were worse than the first portion of the video.

From these settings, we coded each viewers' response on each dimension to indicate whether the enhanced display (+1), the natural display (-1), or neither (0) was preferred. While this method eliminates any magnitude of difference, it allows us to normalize what might otherwise be vastly different magnitude estimations by our subjects. In addition to making judgments on the seven scales after viewing each half of the video, after viewing both portions of the film, subjects were asked which half they preferred (enhanced or natural) and which half appeared «processed».

The observers chose their own viewing distance from the television monitor after watching the opening credits of the movie. Distances ranged from one to six feet with a mean viewing distance of 3.2 feet (mdn =3 ft). While this change in viewing distance affects the retinal spatial frequencies modified by the enhancement algorithm, it is a more reasonable approximation of how visually impaired viewers would use such a system in their homes. Not surprisingly, there was a significant relationship between acuity and distance chosen, such that those viewers with worse acuity sat closer to the screen [$r = -0.55$, $p = 0.014$]. However, there was no relationship between distance (or acuity) and performance in either the natural [$r = 0.20$, $p = 0.414$] or enhanced [$r = 0.28$, $p = 0.244$] conditions.

RESULTS

PERFORMANCE

The percentage of questions answered correctly with no enhancement was 77% (SD = 8.7%). With enhancement, viewers were able to answer 78% (SD = 7.1%) of the questions correctly, demonstrating no difference between the two conditions. However, when we looked at the performance on individual questions, we noted that several (16) of them had been answered correctly by all of the subjects when they watched the video with no enhancement. This is similar to the performance of the normally sighted group in our previous study (24 questions), but much better than the initial low vision group (Peli *et al.*, 1996), who only answered 5 of the questions with 100% accuracy. This improvement in overall performance in the natural condition is likely due to the fact that in our initial study, all viewers sat 6 ft. from the screen, while in the current study, viewers chose their preferred distance (presumably to optimize visual perception).

Because these (16) questions could be answered by all the viewers when the video was shown in the natural condition, there is no room for improvement in the enhanced condition. Therefore, these questions were eliminated and mean performance recalculated. Using the remaining 43 questions, we find a small (5.2%), but statistically significant, improvement in performance when the enhancement was applied to the video signal [76±10% for enhanced; 71±12% for the natural condition; $t(18) = 2.5$, $p = 0.022$]. While this effect is small, it is important to remember that using this same technique, we found only a 15% difference in performance between normally sighted and visually impaired viewers who watched the video with no enhancement (Peli *et al.*, 1996).

Attribute	Subject 1	2	3	5	7	8	9	10	11	12	14	15	16	17	18	19	20	21	22	Mean
Color	-1	-1	-1	0	-1	-1	0	0	1	-1	0	-1	1	-1	-1	-1	1	-1	-1	-0,47*
Details	-1	-1	0	-1	0	-1	0	-1	1	-1	1	1	1	-1	-1	0	1	0	-1	-0,21
Faces	0	-1	0	-1	0	0	0	-1	1	-1	1	1	1	-1	-1	0	1	-1	-1	-0,16
Expressions	0	-1	0	1	0	0	-1	-1	0	1	1	0	1	-1	-1	0	1	-1	-1	-0,11
Story	0	-1	0	1	0	0	-1	0	1	0	0	0	1	0	-1	0	1	0	-1	0.00
Sound	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	0	1	0	-1	0.00
Overall	-1	-1	0	-1	0	-1	-1	0	1	-1	1	-1	1	-1	-1	0	1	-1	-1	-0,37
Sum	-3	-6	-1	-1	-1	-3	-3	-3	5	-3	4	0	7	-5	-7	-1	7	-4	-7	1,32
Enhancement Benefit (%)	2	6	9	31	5	-1	-19	9	8	-14	15	-1	9	6	-9	6	6	-6	2	
Preference	N	N	N		N	N	N	N	E	N	E	N	E	N	N	N	E	N	N	
Processed	N	N	E	N						E	E	E	E	N	E	E		E	E	

* This value was significantly different from 0 [$t(18) = -2.7$, $p < 0.02$].

Table 1. Subjects' preferences for the enhanced or natural conditions for the seven attributes listed, their enhancement benefit and preference after viewing both conditions, and their indication of which condition they thought had been «processed». 1 indicates enhancement was preferred; 0 no preference; - 1 natural preferred. Enhancement benefit is the difference between the enhanced and natural viewing conditions for each subject. Negative numbers indicate better performance in the natural condition. The shaded boxes (E) indicate those viewers who preferred the enhanced condition after viewing both (N indicates preference for natural; a blank box indicates no preference). The same indicators are used in the «processed» row.

Judging performance on the basis of the remaining 43 questions, of the 19 observers, 14 (74%) showed improved performance with the enhancement (mean = 9%; range: 2-31%) and 5 decreased their performance (mean = 5%; range: 1-14%). (See Table 1, *Enhancement Benefit*, for complete data.)

Preference

We asked our subjects to compare the natural to the enhanced conditions on the basis of seven attributes: color, visibility of details, ability to recognize faces, ability to recognize facial expressions, ease of following the story, sound, and overall impression. For each of these attributes weas assigned a value of -1 if the viewer preferred the natural condition, 0 if there was no preference, and + 1 if the enhanced condition was preferred. We also asked them to indicate which condition they preferred overall (after viewing both), and which had been «processed». Their ratings are given in Table 1.

Unlike the viewers in our previous study (Peli *et al.*, 1994b) these subjects did not overwhelmingly prefer the enhanced condition. In fact, only 4 (21%) indicated that they preferred the enhanced portion of the video after viewing both. This was consistent with their coded preference ratings for overall impression. Only these four subjects gave the enhanced portion a higher rating on the «overall» attribute than the natural portion. While few of the subjects indicated a preference for the enhanced condition, there were very few who strongly disliked it. The sum of each subjects' preference ratings indicates the overall strenght of response and can range from -7 (always preferred natural) to +7 (always preferred enhanced). There were only 4 viewers who gave the natural condition a higher rating on 5 or more of the attributes, and the average of each viewers combined ratings (-1.32) was not statistically different from 0 [$t(18) = -13$, $p = 0.20$].

When we look at the individual attributes, only «color» was consistently rated as better in the natural condition (see Table 1). While the enhancement algorithm is applied to the luminance portion of the video signal only (the part seen on a black and white TV), the method of separating the color and recombining it with the enhanced luminance signal results in some color distortion effects. Thus, it is not surprising that many of these viewers found the color appearance more satisfactory in the natural condition.

It is interesting to note that there was little relationship between subjects' performance and their preference for one condition or the other. The correlation between the sum of the preference ratings and the difference in performance between the enhanced and the natural conditions was not significant [$r = 0.32$, $p = 0.19$]. We also looked at the distribution of responses on the «preference» and «processed» questions on the basis of which condition allowed for the better performance for each subject. All of the subjects who preferred the enhanced condition also answered more questions correctly when the video was enhanced. Interestingly, viewers who answered more questions correctly in the enhanced condition were less accurate at indicating which display had been processed, while those who performed better in the natural condition were much more likely to indicate that the enhanced display had, in fact, been processed.

DISCUSSION

The results of the current study are strikingly different from our previous studies. It is therefore important to try to address this difference and the possible causes for it. It is not surprising that all of our subjects did not show an improvement in performance with enhancement. In the original study of the effects of contrast enhancement on face recognition (Peli *et al.*, 1991), only 85% of the subjects showed improvement, and this improvement was statistically significant only for about half of them. It is also not surprising that some of our viewers showed decreased performance with enhancement. When the spatial frequencies critical to face recognition were enhanced, normally-sighted subjects showed decreased performance (Peli *et al.*, 1994a). In the current study, viewers chose their own viewing distance using the natural video image. This should have optimized performance in the natural condition (as evidenced by the large number of questions answered correctly by all of the subjects). Therefore, the enhanced image may have appeared distorted to them, as it did to normally sighted observers in the face recognition task. The variable seating distance and the lack of individual enhancement probably combined to render the effect of the enhancement small or useless for many of the subjects. Even in our previous study, where we found no effect of tuning, we did not reject the possible need for individual tuning (Peli *et al.*, 1994a). In view of the current results, we feel that individual tuning is essential, especially when viewers are free to choose their seating distance from the screen.

Separate from the issue of the type of enhancement applied, we believe that the quantitative method using the AD-based questions is ineffective for this investigation. The large number of questions answered correctly without enhancement limits the range of possible improvement, and the fact that even normally-sighted observers are unable to answer all of the questions correctly also reduces the range of possible improvement. It is not clear if the fault is with the questions we formulated from the AD script or in the items described. Although we remain convinced that AD is a very useful tool for blind audiences, we are not convinced that the visual information contained therein, which is of some use for partially sighted audiences (as we have shown) can be directly used to test performance in the widely divergent low vision population. Our experience with this approach suggests that performance evaluation from motion video remains elusive due to the vast complexity of the task itself and the confounding effects of other factors, such as attention, prior knowledge, and objects of interest.

The preference data gathered from this study are quite different from our initial study where 95% of the viewers preferred the enhanced display. In the current study, only 21% of the subjects indicated a preference for enhancement, although they were very clear about their strong degree of preference. As with the performance measure, this may have been due to the application of standard enhancement settings for all subjects. The freedom to choose viewing distance, combined with standard enhancement settings, would result in ideal enhancement for some viewers (in terms of which retinal spatial frequencies are enhanced), and a processed image that looks distorted (as it does to normally sighted viewers) to others. In future investigations, we will concentrate on assessing preference, over time, in response to changes in enhancement parameters. Such continuous assessment of perceived image quality has been demonstrated for normally sighted observers (Hamberg & de Ridder, 1995) and may be useful in both selecting enhancement parameters and assessing preference between enhanced and unenhanced videos.

Acknowledgments

This work was supported by NIH grant #EY05957 to E. Peli. E. Fine was supported by grant #EY06632 during preparation of this manuscript. We thank Julie Irish and Angela Labianca for help in developing the questions used in this study, as well as Barry Cronin and Gerry Field of WGBH, Boston, for providing us with program videos, software, and valuable technical help.

REFERENCES

BERKOWITZ, M *and cols.*: (1979). *Characteristics, Activities and Needs of People with Limitation in Reading Print.* NY: American Foundation for the Blind.

CRONIN, B. J., and KING, S. R.: (1990, december). The development of descriptive video service. *Journal of Visual Impairment and Blindness,* 503-506.

FINE, E. M., and PELI, E.: (1995). Enhancement of text for the visually impaired. *Journal of the Optical Society of America A, 12* (7), 1439-1447.

HAMBERG, R., and de RIDDER, H.: (1995). Continuous assessment of perceptual image quality. *Journal of the Optical Society of America A, 12* (12), 2573-2577.

JOSEPHSON, E.: (1968). *The Spatial Life of Blind People* (Research Series 19). NY: American Foundation for the Blind.

PELI, E., AREND, L. E., Jr., and TIMBERLAKE, G. T.: (1986, september). Computerized image enhancement for visually impaired persons: New technology, new possibilities. *Journal of Visual Impairment and Blindness,* 849-854.

PELI, E., FINE, E. M., and LABIANCA, A. T.: (1996). Evaluating visual information provided by audio description (AD). *Journal of Visual Impairment and Blindness,* in press.

PELI, E., FINE, E. M., and PISANO, K.: (1994b). Video enhancement of text and movies for the visually impaired. In A.C. Kooijman, P.L. Looijestijn, J.A Welling and G.J. van der Wildt (Eds.), *Low Vision: Research and New Developments in Rehabilitation* (pp. 191-198). Amsterdam: IOS Press.

PELI, E., GOLDSTEIN, R. B., YOUNG, G. M., TREMPE, C. L. and BUZNEY, S. M.: (1991). Image enhancement for the visually impaired. *Investigative Ophthalmology and Visual Science, 32* (8), 2337-2350.

PELI, E., LEE, E., TREMPE, C. L., and BUZNEY S.: (1994a). Image enhancement for the visually impaired: The effects of enhancement on face recognition. *Journal of the Optical Society of America A, 11* (7), 1929-1939.

PELI, E., and PELI, T.: (1984). Image enhancement for the visually impaired. *Optical Engineering, 23* (1), 47-51.

PELI, T., and LIM, J. S.: (1982). Adaptive filtering for image enhancement. *Optical Engineering, 21,* 108-112.