

The Impact of Field of View on Understanding of a Movie Is Reduced by Magnifying Around the Center of Interest

Francisco M. Costela^{1,2} and Russell L. Woods^{1,2}

¹ Schepens Eye Research Institute, Massachusetts Eye and Ear, Boston, MA, USA

² Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Correspondence: Russell L. Woods, Schepens Eye Research Institute, 20 Staniford St, Boston, MA 02114, USA. e-mail:

russell_woods@meei.harvard.edu

Received: December 4, 2019

Accepted: May 12, 2020

Published: July 7, 2020

Keywords: Magnification; center of interest; zoom; field of view; movies

Citation: Costela FM, Woods RL. The impact of field of view on understanding of a movie is reduced by magnifying around the center of interest. *Trans Vis Sci Tech.* 2020;9(8):6, <https://doi.org/10.1167/tvst.9.8.6>

Purpose: Magnification is commonly used to reduce the impact of impaired central vision. However, magnification limits the field of view (FoV) which may make it difficult to follow the story. Most people with normal vision look in about the same place at about the same time, the center of interest (COI), when watching “Hollywood” movies. We hypothesized that if the FoV was centered at the COI, then this view would provide more useful information than either the original image center or an unrelated view location (the COI locations from a different video clip) as the FoV reduced.

Methods: The FoV was varied between 100% (original) and 3%. To measure video comprehension as the FoV reduced, subjects described 30-second video clips in response to two open-ended questions. A computational, natural-language approach was used to provide an information acquisition (IA) score.

Results: The IA scores reduced as the FoV decreased. When the FoV was around the COI, subjects were better able to understand the content of the video clips (higher IA scores) as the FoV decreased than the other conditions. Thus, magnification around the COI may serve as a better video enhancement approach than simple magnification of the image center.

Conclusions: These results have implications for future image processing and scene viewing, which may help people with central vision loss view directed dynamic visual content (“Hollywood” movies).

Translational Relevance: Our results are promising for the use of magnification around the COI as a vision rehabilitation aid for people with central vision loss.

Introduction

Video content, displayed on television, in movies, and on the internet, is a major source of information, entertainment, and social engagement.^{1–3} Its importance is demonstrated by how, despite a reduced viewing experience due to vision impairment, on average, people with central vision loss (CVL) watch at least as much television (TV) as people with normal sight.⁴ That is even though they express dissatisfaction with their viewing experience⁴ and have an impaired ability to follow the story.⁵

Magnification is the most common and an effective form of visual aid for CVL, provided through relative-size and relative-distance magnification, and

instruments and devices such as optical and electronic handheld magnifiers, bioptic telescopes, closed-circuit-television devices, and electro-optical head-mounted displays. Currently, rehabilitation for TV viewing is very limited for people with CVL. Overall, the benefits found for video viewing with contrast enhancement,^{6–9} and edge enhancement of video,^{10,11} have been modest and no commercial device has been available apart from the Belkin DigiVision DV1000 device (that was marketed to people with normal vision).⁶

Magnification using devices and instruments necessarily restricts the amount of information visible in the field of view (FoV), through the interaction between the magnification and the extent of the display or exit pupil of the instrument or device. This can cause a loss of information and context, and diminish the

viewing experience, despite the ability to resolve details that would not have been visible but for the magnification. For example, when the viewing area is fixed, as with a monitor or other display, with $2\times$ magnification, the FoV contains 25% (1/4) of the original image, and with $6\times$ magnification, the FoV is only 2.8% (1/36). Such a reduction in the amount of information available may lead to substantial changes in information acquired or visual task performance. Spatial awareness is impaired by restricting the FoV¹² and small visual fields.¹³ Similarly, pedestrian mobility is impaired by FoV restriction^{14,15} and small visual fields.^{16,17} Restricted peripheral vision, through FoV restriction^{18,19} and visual field loss,^{20,21} is also related to worse driving performance.

Although peripheral vision has many limitations as compared to foveal vision, including local ambiguity of the location and phase of features,²² the gist of a scene can be obtained quickly from peripheral vision.^{23,24} Although the ability to perform vision-related tasks decreases as FoV reduces,^{12–21} the amount or proportion of visual content necessary for recognition or comprehension of visual content is not clear. In a clever and evocative study, Ullman et al.²⁵ quantified the transition in recognition rate from a minimal recognizable configuration (MIRC) image to a nonrecognizable descendant (by sequentially cropping 20% of the image). This reduction in recognition rate was quantified by measuring a recognition gradient, defined as the maximal difference in recognition rate between the MIRC and its five descendants. The average gradient was 0.57 ± 0.11 , suggesting that small changes at the MIRC level can make the picture unrecognizable. These results, found with static images, raise an interesting question regarding the importance of peripheral and contextual information in dynamic settings, and if people are in fact able to understand visual information when only a subset of it is displayed. In our study, we begin to address this issue in video by showing restricted views. We hypothesized that there would be a reduction in video comprehension as FoV size reduced.

While viewing “Hollywood” movies (video in which the content was directed²⁶), people with normal vision look in about the same place most of the time.^{26,27} We assume that this between-viewer consistency is because there is often a characteristic of the scene (e.g. a close-up image of a face, a full moon in an empty sky, or a brightly-colored bird on a branch) that draws near universal attention. We termed this area the center of interest (COI). The series of COIs (one per frame) within a video clip is the “democratic” video scan path. We presume that most of the information necessary to follow the story is contained in the democratic-COI scan path, as the director of the video has designed the

scene to draw the viewer’s gaze to particular locations, the COIs. A small FoV might not include the COI.

We hypothesized that if the FoV location was centered around the democratic COI, then this area would provide more useful information than simply centering the FoV around the original image center, as happens with simple magnification. However, it should be noted that the COI is often in the middle of the original image,²⁷ which may limit the value of the COI as the FoV-center. As a control condition, we included FoV-center around an unrelated view location, defined by the COI of a different video clip (i.e. similar characteristics but not related to the content). We varied the FoV-size related to magnification as used to assist people with CVL, and used a recently-described, objective technique to measure the ability to follow the story. We hypothesized that the dynamic aspects of video clips may ameliorate the impact of the restricted FoV, as compared with the drastic effects on recognition reported by Ullman et al.²⁵ who restricted the FoV of static images. Our study may have implications for the development of new methods to modify dynamic electronic images (videos as in TV or movies) to assist people with CVL.

Methods

Subjects watched and then described twenty 30-second video clips that varied in the FoV of the original content that was visible (amount of available information) and in the manner in which the FoVs were selected from the original image (i.e. the locations of the subsets of visual information). The ability to follow the story was measured using the sensory information acquisition (IA) method.²⁸ The study involved 3 groups of subjects comprising 60, 432, and 128 subjects.

Experimental Conditions

For each of the 20 video clips, we created new versions that contained 50%, 25%, 11%, 6%, 4%, or 3% of the original scene (see Figs. 1b–e). When those FoVs were expanded to the original size of the video clip, effectively, they provided 1.4, 2, 3, 4, 5, and 6 times magnification, respectively. These magnifications (and thus FoV sizes) are in the range of prescribed devices clinically. In total, there were seven FoV sizes of each video clip, one unrestricted-area condition (100%), and six reduced FoV conditions (50% to 3%). For the six reduced FoV-size conditions, there were three different FoV-center conditions that were around: (1) the original image center; (2) the democratic COI (deter-

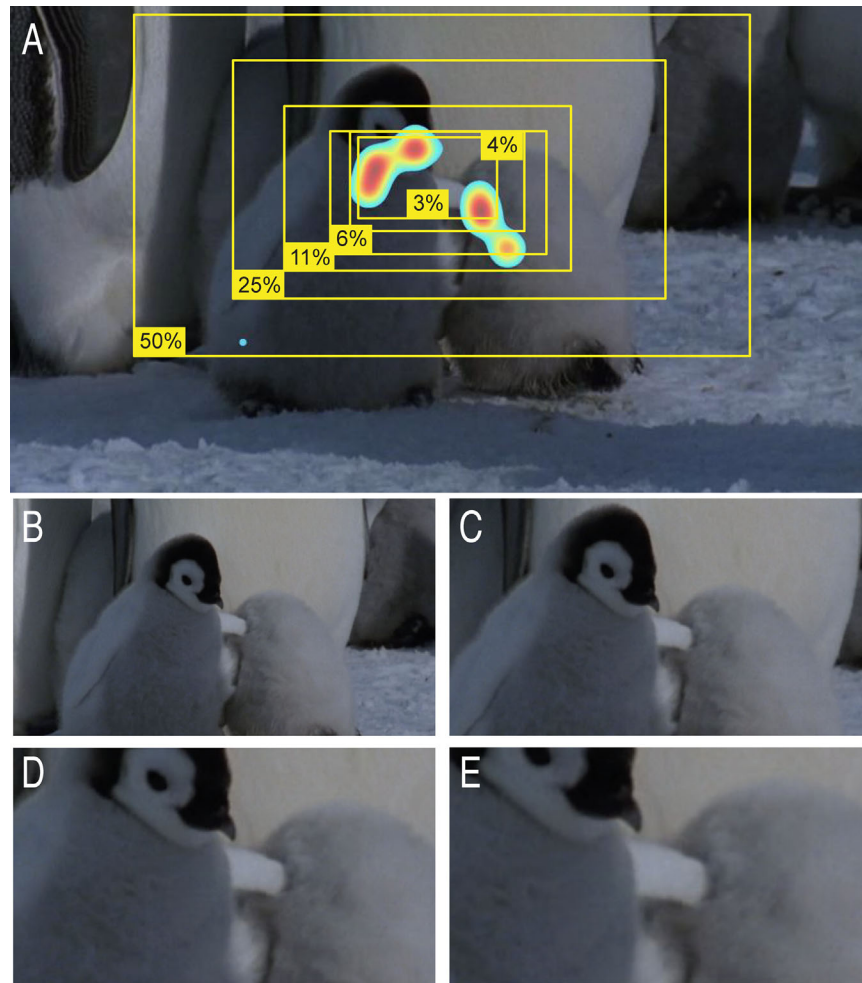


Figure 1. (a) Original frame with gaze density kernel and six field of view (FoV) boxes. The color map indicates the kernel density estimate of gaze positions from group 1 subjects for this frame. *Yellow rectangles* represent the FoV boxes computed from the democratic COI for FOVs of 50%, 25%, 11%, 6%, 4%, and 3%. FoVs boxes enlarged to original screen size are shown for (b) 50%, (c) 25%, (d) 11%, and (e) 6% FOVs of that frame. *Blue dot* in lower left corner within the 50% box corresponds to a gaze point.

mined as described below); and (3) an unrelated COI (from a different video clip). Thus, there were a total of 19 conditions ($1 + 3 \times 6$). The first FoV-center condition (#1) represented simple magnification, as has been supplied with some video-viewing devices. The unrelated-video FoV-center condition (#3) was a control condition. For that, the COI was derived from a different video clip by randomizing the order of the 20 clips so that unrelated gaze data were used to compute the COI for every clip.

It is possible that subjects would be able to maintain understanding because they had the audio track available. Previously, by reviewing the responses, we had found that there was very little information related to the audio content,²⁸ but had not formally tested the effect of audio content. We hypothesized that any benefit from audio would be greatest at small FoVs,

when the audio track might provide some context that improved the description, and thus increased the IA score. So, we implemented four extra conditions to test whether subjects were using audio information to follow the story and thus improving their description, despite instructions to only report visual information. For this control condition, we removed the audio information from the original viewing condition (100%) and from the three area-center conditions with a FoV of 3%. This added 4 experimental conditions, for a total of 23 experimental conditions in the study.

Subjects and Their Tasks

There were three groups of subjects involved in this study. The first group consisted of 60 subjects who watched the video clips in the laboratory (lab sourced)

Table. Self-Reported Demographic Characteristics of Subjects

		Lab-sourced Group 1 & 3 (N = 60)	Crowd-sourced Group 2 (N = 432)	Crowd-sourced Group 3 (N = 68)
Gender	Male	30 (50%)	228 (53%)	46 (68%)
	Female	30 (50%)	204 (47%)	22 (32%)
Age (median, min-max)		64 y (23-85 y)	31 y (18-69 y)	38 y (22-66 y)
Race	Black	5 (8%)	28 (6%)	4 (6%)
	White	54 (90%)	341 (75%)	56 (82%)
	Asian	1 (2%)	36 (8%)	2 (3%)
	American Indian/Alaska native	0 (0%)	3 (1%)	1 (1%)
Ethnicity	Multiple	0 (0%)	24 (5%)	5 (8%)
	Hispanic	1 (2%)	43 (10%)	6 (8%)
	Not Hispanic	59 (98%)	363 (84%)	62 (92%)
	Not reported		26 (6%)	
Highest education	High school	5 (8%)	51 (12%)	8 (12%)
	Some college	6 (10%)	122 (28%)	9 (13%)
	Associate degree	2 (3%)	126 (29%)	22 (32%)
	Bachelor's degree	20 (33%)	106 (24%)	20 (30%)
	Postgraduate degree	27 (45%)	27 (6%)	9 (13%)

There were missing data for age, gender, and education for 23 of the subjects in group 2 (crowdsourced).

who have been described before.^{5,29} We used their gaze (eye movement) data to determine the democratic COI. This group is described in more detail below and in the [Table](#). The second group consisted of 432 crowd-sourced subjects who viewed at least one of the 23 experimental conditions. More detail about this group is provided below and in the [Table](#). The third group consisted of the 60 lab-sourced subjects (equal to the first group) and 68 crowd-sourced subjects, who provided descriptions of the video clips in their original format. Their responses formed the control (or “crowd”) database of responses that were used for scoring the responses of the group-2 subjects, as described below. They have been described previously.^{28,29}

Group 1 – Gaze-Tracked While Viewing Unrestricted (Original) Video Clips

Lab-sourced subjects were recruited from the community in and near Boston, Massachusetts, equally for three age strata: under 60 years, 60 to 70 years, and > 70 years, each with equal numbers of men and women. The demographics are presented in the [Table](#) and details about eligibility criteria have been previously reported. Each lab-sourced subject wore their habitual, not necessarily optimal, optical correction while viewing the original video clips on a 27” diagonal 16:9 aspect ratio display at 100 cm. The videos were all

33° wide, but had variable height (up to 19°) depending on the aspect ratio of the original material. The clips were displayed using a MATLAB program using the Psychophysics Toolbox³⁰ and Video Toolbox.³¹ Subjects’ head movements were restrained with a head and chin rest for the duration of the experiment. An SR Research EyeLink 1000 infrared eye tracking system was used to collect gaze (eye movement) data during video clip presentations. For each of the 20 video clips, we used these data to determine the democratic COIs for each clip (see COI determination below).

Group 2 – Viewed and Described Unrestricted and Restricted-Area Video Clips

Crowd-sourced subjects were recruited through postings on Amazon Mechanical Turk and were limited to workers who were registered as living in the United States.²⁹ Demographic information, including gender, race, age, education level, and TV watching habits - number of hours watching TV (7 ordered categories; from 0 hours to over 5 hours a day) and reported difficulty (five ordered categories; from never to always), was requested from each worker before they completed any tasks. At the end of the demographic survey, workers were informed about what they would be asked to do and actively consented by selecting a check box. These workers were anonymous, known to us only by an ID assigned by Amazon. They were

paid on a per-response basis, with Amazon as an intermediary. Workers were paid US \$0.25 per response contributed, with a one-time \$0.25 bonus for filling out the demographic survey and a \$0.25 bonus for every 10 responses contributed and approved. A total of 432 subjects viewed the edited video clips (all 23 conditions; see experimental conditions below) within a Web browser, on a local computer of their choice. Therefore, the size of the monitor, their distance from the monitor, and other display characteristics were not fixed and not known to us. The clips were shown within the frame of the Mechanical Turk interface, with each clip representing a separate Human Interface Task (HIT; the unit of paid work on the Mechanical Turk website). Below the clip, there were two video description prompts to input text into boxes (described below). Text entry into these boxes was disabled until the video clip had finished playing. Workers could complete as many video clip description tasks (HITs) as they wanted while more clips that they had not seen were available, at any time of day. It was not possible to guarantee that each worker would complete a certain number of these tasks. Workers were prevented from seeing any clip more than once. Across all crowdsourced subjects, 125 to 156 responses were collected for each experimental condition, for a total of 3,334 responses.

Data collection for the crowd-sourced responses were contributed by 432 distinct Mechanical Turk worker IDs, (median age = 31, range = 18-69 years) during 29 days of active data collection. The median number of responses contributed by crowdsourced subjects was 7, range 1 to 20. Responses were often contributed over the course of multiple working sessions.

Group 3 – Control Group Viewed and Described Unrestricted (Original) Video Clips

As described previously,²⁸ 60 lab-sourced subjects (who also had their gaze tracked; group 1) and 68 crowd-sourced subjects provided descriptions of the video clips in their original format.

Comparing the Three Groups of Subjects

The demographics of the three samples are presented in the Table. The lab-sourced sample was older than crowdsourced group 3 (Wilcoxon rank-sum test, $z = 7.00$; $P < 0.001$), which was older than crowdsourced group 2 ($z = 3.59$; $P < 0.001$). There was a higher proportion of white subjects in all groups than found in the general population in the United States. None of the lab-sourced sample reported their ethnicity as “multiple,” in contrast to approximately 7% of the crowdsourced samples. Race tended toward

the lab-sourced group having a higher proportion of people reporting their race as white and fewer reporting Asian than crowdsourced group 2 ($X^2(3) = 7.94$; $P = 0.05$). The lab-sourced sample had a high proportion of people with postgraduate degrees. The distributions of education levels differed between group 1 and group 2 (Kolmogorov–Smirnov test, $D = 0.47$; $P < 0.001$) and group 3 ($D = 0.36$; $P = 0.001$). Gender did not vary significantly between groups 1 and 2 ($X^2[1] = 0.21$; $P = 0.65$) but group 3 tended to have a higher proportion of men than group 2 ($X^2(1) = 4.99$; $P = 0.03$) or group 1 ($X^2[1] = 4.12$; $P = 0.04$). Age, gender, and education were included as covariates in analyses of the IA scores.

Information Acquisition Measurement

A natural-language approach was used to determine the IA score. Following each 30-second video clip, the viewer was given the prompts: “Describe this movie clip in a few sentences as if to someone who hasn’t seen it” and “List several additional visual details that you might not mention in describing the clip to someone who hasn’t seen it.” This measurement method has been reported in detail previously.²⁸ In summary, the database of responses provided by subjects in group 3 were used to compute the information acquisition measurement from responses in group 2. For each response about a video clip by each subject in group 2, the response was compared, one by one, to each response about that video clip in the control database (made by subjects in group 3 who saw the original, 100%, clip version). In each paired comparison, the number of shared words was counted. The IA score for each video clip for each subject was the average of the shared-word counts (after removing stopwords) and disregarding repeated instances of the word in either response.

Democratic COI Determination

Each subject in group 1 watched 10 to 13 of the 20 clips once. Subject’s gaze was tracked at 1,000 Hz. Video frames were shown at 30 Hz, so each subject could contribute up to 33 data points per frame. Saccades were removed from the data. For each video frame of each clip, the remaining data (fixations and pursuits) for all subjects who viewed that frame were used to compute a kernel density estimate. To determine the democratic COI of each frame, we integrated the area under the region of the density estimate for all possible positions of a restricted-area box over the frame, using a symmetrical Gaussian function. The restricted-area box had the same aspect ratio as the

original clip altered in size so as to contain the required proportion of original frame. For example, an FoV box of 25% had sides that were $\frac{1}{2}$ the width and $\frac{1}{2}$ the height of the original video frame. The democratic COI for that FoV was defined as the center of the FoV box with the highest integral value. That process was repeated for each frame of each video clip for each FoV box size. The rationale for using the integral of the gaze-density distribution was that it accounts for multimodal distributions better than taking an average or median of the gaze locations. Figure 1A shows an example of the FoV boxes, computed for 50%, 25%, 11%, 6%, 4%, and 3% of the original scene, superimposed over the original frame. Once the democratic COI coordinates were obtained, to avoid jitter from small changes in the gaze-density distributions between frames, we applied a deadband filter of 60 pixels followed by a smooth quadratic filter with a span of 10% of the data (temporal smoothing of democratic COI location). Then, for each frame, we rescaled every FoV box, centered at the COI, to the original clip dimensions. That is, we magnified by the inverse of the FoV (e.g. if 6% FoV, then it was magnified 4 \times). Figures 1B to 1E show the rescaled box for four of the FoVs.

Video Clips

There were twenty 30-second video clips, chosen to represent a range of genres and types of depicted activities. The genres included nature documentaries (e.g. BBC's Deep Blue, The March of the Penguins), cartoons (e.g. Shrek, Mulan), and dramas (e.g. Shakespeare in Love, Pay it Forward). The clips included conversation, indoor and outdoor scenes, action sequences, and wordless scenes in which the relevant content was primarily the facial expressions and body language of one or more actors.

We conducted a post hoc rating of video content, described previously.³² In summary, each video clip was categorized for: (1) Number of cuts (low [< 4], medium [4 to 5], or high [> 5]); (2) lighting (low, medium, or high); (3) environment (indoor or outdoor); (4) auditory information (low, medium, or high), and the importance (low, medium, or high) of each of (5) faces, (6) human figures, (7) man-made objects, and (8) nature for understanding of the video content.

The process for creating the frames for the other two FoV-center conditions was similar to the process of creating the democratic COI video frames. For the center FoV condition, the FoV boxes were always centered on the center of the original frame. For the control condition that used unrelated view locations, we used the democratic COI locations from a differ-

ent clip. For that, each video clip ("A") was randomly paired with another clip ("B"). Then, the restricted-area centers found for clip B (including temporal smoothing) were applied to create the FoV boxes for clip A. For all three FoV center conditions, the FoV boxes were expanded by the required magnification to return the frame to the original frame size. Finally, each FoV size and FoV center condition video was reconstructed from the constituent frames. So, all experimental video clips had the size and aspect ratio of the original clip (see Fig. 1).

Statistical Analyses

To examine the effects of FoV center and FoV size, we used a mixed-effects model (also known as a linear mixed model) with FoV size as a continuous variable, and an interaction between the fixed factors FOV center and FoV size, with age, education, and gender as covariates, and subject and video clip as fully crossed random factors.³³ FoV-size was implemented as the logarithm (base 10) of the FoV (visible area), as this produced the most parsimonious extrapolation of the IA score reaching a value of zero at some small FoV size. To complete the model structure, we randomly (arbitrarily) assigned trials with 100% visible area to one of the three FoV center categories. The model was constrained so that the fits (curves) for each FoV center passed through the same IA score value at 100% FoV size, because there is no reason that they should differ. Thus, the fits for each FOV center condition could only differ in slope.

To examine whether subject-dependent factors were related to IA scores, race and the amount of TV watched and the difficulty watching TV reported by the subjects were added to the main model, as described above, that already included age, gender, and education. Then, all of the subject-dependent factors that were not significant ($P > 0.10$), were sequentially removed from the model. Then, to examine the effects of video-dependent factors (e.g. importance of faces for understanding), all eight video-dependent variables were added to the model, and then were sequentially removed from the model if the variable was not significant ($P > 0.10$).

To examine the effects of auditory information on IA scores, we used a different mixed-effects model with auditory track presence and FoV center as fixed factors, with age, education, and gender as covariates, and subject and video clip as fully crossed random factors. In all analyses, we accepted $P \leq 0.01$ as significant and $0.01 < P \leq 0.10$ as a "trend."

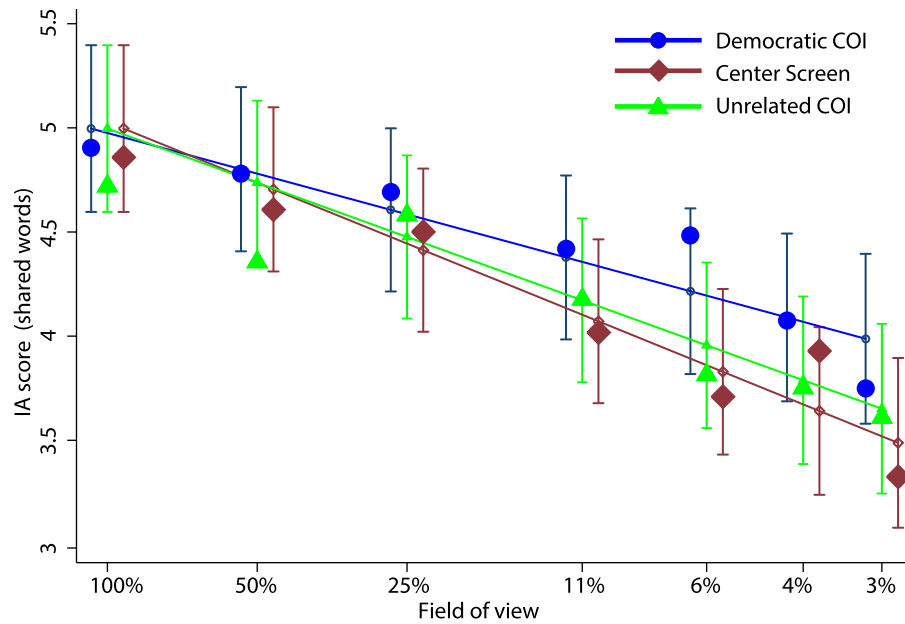


Figure 2. Effects of FoV and viewing condition on IA score, for FoVs centered on the democratic COI (blue circles), an unrelated COI (light green triangles), and at the center of the screen (dark-red diamonds). The solid lines and small symbols represent the fit. Error bars indicate 95% confidence intervals of the fit. Filled shapes represent the average IA score of all subjects for that condition, corrected for clip and subject.

Results

As we hypothesized, overall (across the three FoV center conditions), IA scores (ability to follow the story of the video clip) reduced as the FoV became smaller ($B = 0.83$; 95% confidence interval [CI] = 0.72 to 0.93; $z = 15.3$; $P < 0.001$). In addition, as we hypothesized, when the FoV center was the democratic COI, the reductions in IA scores were less with increasing restriction of the visible area (i.e. shallower slope; $B = 0.65$; 95% CI = 0.52 to 0.77; $z = 10.11$; $P < 0.001$) as compared with the FoV center being the original image center ($\Delta B = 0.32$; 95% CI = 0.20 to 0.44; $z = 5.36$; $P < 0.001$) or an unrelated view location (COI of a different video clip: $\Delta B = 0.22$; 95% CI = 0.10 to 0.33; $z = 3.60$; $P < 0.001$), as shown in Figure 2. The change in IA score with reducing visible area tended to be less with the unrelated center than the original image-center ($\chi^2[1] = 3.14$; $P = 0.08$).

Effects of Subject and Video Characteristics

The reported number of hours watching TV and difficulty watching TV were not related to age, gender, education, or race, except for a trend for hours watching TV to decrease with increasing age (ordered logistic regression; $z = 1.80$; $P = 0.07$) and with increasing education ($B = -0.19$; 95% CI = -0.38 to -0.002; $z = 1.98$; $P = 0.05$). In the backward stepwise, mixed-effects

regression of subject-dependent factors, race, number of hours watching TV, and difficulty watching TV were not related to IA scores, so were removed. Men had a lower IA score than women by 0.50 shared words (95% CI = -0.60 to -0.39; $z = 9.61$; $P < 0.0001$), IA score reduced with increasing age by 0.21 shared words per decade ($B = 0.021$; 95% CI = 0.01 to 0.02; $z = 7.62$; $P < 0.001$) and increased with increasing education level ($B = 0.04$; 95% CI = 0.02 to 0.07; $z = 3.27$; $P = 0.001$).

The video-dependent factors – the importance of faces, human figures, man-made object and nature for understanding the clip, and number of cuts, lighting, environment, and auditory information – were unrelated, except that nature importance was related to environment (Spearman rho = 0.62; $P = 0.004$), and there were trends for nature importance to increase (rho = 0.41; $P = 0.08$) and audio information to decrease (rho = -0.49; $P = 0.03$) with increasing lighting, and for nature importance to decrease with increasing face importance (rho = 0.47; $P = 0.04$), in these 20 video clips. To the model just developed (that included age, gender, and education), we added all of the video-dependent factors and conducted another backward regression. Indoor scenes tended to have higher IA scores than outdoor scenes by 0.69 shared words (95% CI = -1.23 to -0.102; $z = 2.30$; $P = 0.02$). IA scores tended to decrease with increasing importance of nature ($B = -0.09$; 95% CI = -0.21 to 0.01; $z = 1.86$; $P = 0.06$) and tended to increase with increasing importance of man-made objects ($B = 0.24$; 95%

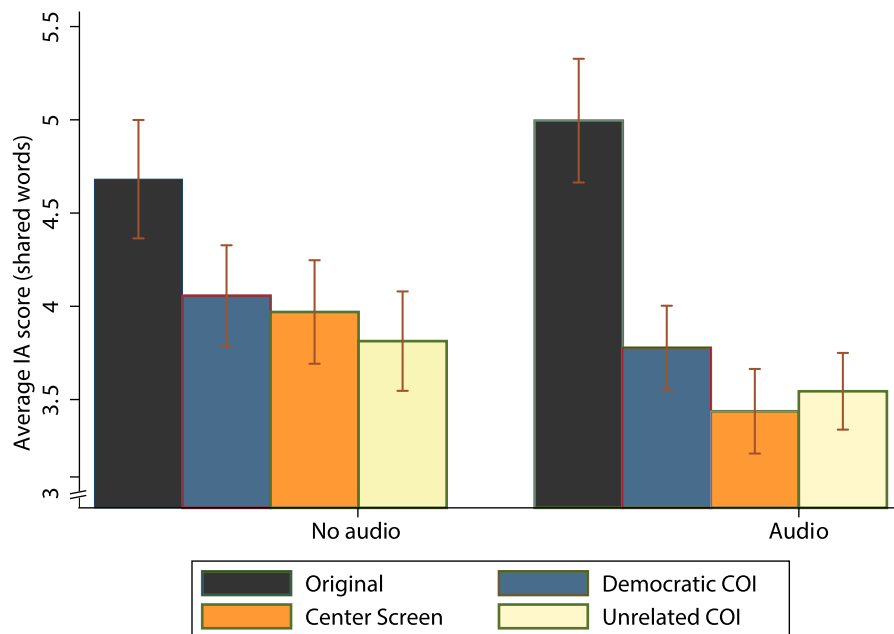


Figure 3. Effect of audio on IA score. Mean number of words shared with responses to the same clip in the crowdsourced dataset, with original clips (black columns) and when viewing 3% of the original image was centered (1) around the democratic COI (blue), (2) the original center of the screen (orange), and (3) on an unrelated COI (yellow). Error bars indicated 95% confidence intervals.

CI = 0.04 to 0.43; $z = 2.39$; $P = 0.02$). In an in-person study, Reeves et al.³⁴ found similar effects, with man-made object importance increasing, and nature importance decreasing IA scores. The other content-related factors (importance of faces, human figures, or auditory content, or the number of cuts per clip, or lighting level) were not significant ($P > 0.10$) so they were removed from the model.

Effect of Audio on IA Scores

In the primary study (reported above), subjects heard the original audio track, but were instructed to report only on the visual aspects of the clip, regardless of audio content. However, subjects may have used the audio content and thereby improved their performance. We hypothesized that if there was benefit of the audio track for clip understanding, that would be greatest for the two FoV center conditions that were less likely to include the democratic COI, the original image and unrelated clip FoV center conditions and would be most pronounced for at the smallest FoV size (3%). First, we examined the effect on the original (100%) viewing condition, and found no difference between the audio-on and audio-off conditions ($z = 0.65$; $P = 0.53$), when corrected for age, gender, and education (Fig. 3). For the 3% FoV size, there was a trend for a reduction in IA scores with audio, by 0.46 shared words ($X^2[1] = 3.78$; $P = 0.05$), and no differ-

ence between the conditions in the effects of audio on IA scores ($z \leq 0.57$; $P \geq 0.57$), when corrected for age, gender, and education. Thus, the subjects were not using audio content to follow the story (or the audio tended to have a negative effect). This result indirectly confirmed that the responses contained in our control (crowd) database of responses (group 3) were using visual rather than auditory cues.

Discussion

Reduced visual field or FoV extent is associated with decreased spatial awareness,^{12,13} pedestrian mobility,^{14–17} and driving,^{18–21} presumably because the available information is reduced. A major impact of the restricted FoV is the loss of peripheral information, where peripheral was a function of the FoV center. Peripheral vision provides scene gist^{23,24} and guides eye movements that direct the gaze to new objects of interest.²² We predicted a reduction in performance (IA scores) as a function of FoV size. Consistent with our hypothesis, we found the expected reduction in performance (IA scores) with reducing FoV size (see Fig. 2). However, we were surprised by how well the subjects could understand and describe the video content with the smaller FoVs. For example, even with only 3% of the original scene available, the IA score was reduced

by only about 1.0 shared words as compared to the unrestricted (100%) view from, on average, from 4.8 to 3.8 shared words. These results show that people can still follow much of the story with a substantially reduced FoV.

This study builds upon the study by Ullman et al.²⁵ by extending their work from static images to video. They quantified the minimum amount of information required to recognize the class (category) of the primary object within the image. To achieve that, they systematically reduced that FoV, then magnified the FoV up to a standard image size. Our approach was similar, except that we did not reduce the FoV to such small sizes and we compared subjects' descriptions to a control database of descriptions, so they made no assumptions about video content. Video comprehension is a much more complex task than categorization. We found a decrease in IA scores with a reduction of the available scene (FoV size); but this reduction was much less dramatic than found by Ullman et al.²⁵ This difference is almost certainly because we did not reduce the FoV sufficiently (small enough). We did not use smaller FoV sizes, as the magnification associated with the 3% FoV of $6\times$ is the largest magnification that is likely to be used by people with CVL when watching videos. The dynamic aspects of video, even at the smallest FoV sizes that we used, seemed to allow the viewer to identify features of both foreground and surrounding objects, as objects that may be included in the description moved in and out of the FoV, and thus minimized the impact of FoV restriction as compared to what might have occurred with a static image.

We asked whether, when the FoV is restricted, there is an advantage to presenting around the democratic COI for acquisition of visual information as compared with two other approaches for determining the location of the viewing area (FoV). We found that the democratic COI approach outperformed the other two approaches (see Fig. 2) as the FoV decreased. This is consistent with our expectations, as we anticipated that there would be little to no effect with the larger FoVs, as there would be substantial overlap of the FoVs between FoV center conditions due to the size of the FoVs. As the FoV decreased, there would have been less frequent overlap between the FoV center conditions, even though the COI is in the center of the original image a large proportion of the time in videos.²⁷ Tseng et al.³⁵ showed a similar center bias of photographers to place structured and interesting objects in the center of the photograph (static image). For video, Goldstein et al.²⁷ found that 73% of COIs were outside the central 4% of the original image area, and 50% were outside the central 6.25% of the original image area. Thus, at smaller FoV sizes, there is

opportunity for the screen center and unrelated area center approaches to miss objects of interest. However, that effect may be countered by motion within the video, which might explain the small differences in the amount of information obtained between the FoV center approaches.

Magnification for TV or movies can be provided with bioptic and spectacle-mounted telescopes, although there is little evidence of their effectiveness, and they are dispensed infrequently to people with CVL.^{4,36} Head-mounted electro-optical devices, including mounted smart phones, have been reported to be used for viewing TV by people with CVL, and can provide magnification, local ("bubble") magnification, and contrast enhancement.³⁷ The impact of the FoV on these devices is unknown. A smaller FoV is associated with slower reading rates,^{38–40} although reading rates may not decline dramatically until the FoV becomes very small.^{38,39}

There was no effect of audio track presence on IA scores when viewing the original (100%) clips, and a trend for a reduction in IA scores with the audio track present when viewing the smallest FoV condition, 3%. We had hypothesized that if audio was being used to follow the story (and thus increasing the IA score), then the effect would be strongest in the 3% FoV center conditions that were not the democratic COI, and thus would often not include the object of interest. We did not find that. Instead, IA scores were, on average, 0.5 shared words lower with the audio track than without, and there was no difference between the FoV center conditions. We speculate that audio could act as a distractor when viewing a restricted FoV. Indirectly, this result confirmed the robustness of our control database of responses, in that, responses were mostly based on visual, and not auditory, cues.

Our results are promising for the use of magnification around the democratic COI to provide a visual aid (vision rehabilitation) to help people with CVL watch videos. As the view area decreased (i.e. magnification increased), centering around the democratic COI reduced the effects of the restricted viewing area compared to simply centering around the original image center. This approach may have the added benefit for people with CVL that it should reduce the need for eye movements to locate objects of interest because the object of interest is always in the center of the magnified (and restricted FoV) view. We anticipate that subjects with CVL will benefit directly from watching video that has been enhanced with magnification around the COI. We plan future studies to determine whether magnification (which reduces the impacts of reduced resolution) around the COI effectively increases the IA scores (as compared to the

original view) in subjects with CVL, who benefit from magnification due to their reduced resolution. An extension to that approach might be to give the viewer the ability to control the presence of the magnification, as has been developed for use with reading⁴¹ and face recognition.⁴² Developing vision rehabilitation methods that modify electronic dynamic images (e.g. TV, movies, and internet videos) to assist people with CVL is worthy of future work, as people with CVL report difficulties watching video⁴ and have reduced ability to follow the story.⁵

Acknowledgments

Eli Peli conceived of the approach of magnification around the COI as a method to assist people with central vision loss. The authors thank Daniel R. Saunders and John F. Ackermann for technical assistance.

Supported by National Institutes of Health (NIH) awards R01EY019100 and P30EY003790.

Disclosure: **F.M. Costela**, None; **R.L. Woods**, None

References

1. Kubey R, Csikszentmihalyi M. Television addiction is no mere metaphor. *Sci Am*. 2002;286:74–80.
2. McQuail D. *Mass Communication Theory: An Introduction* 2nd edn. London: Sage; 1987.
3. ThinkBox. Screen Life: TV in demand. Available at: <http://www.thinkboxtv/Research/Thinkbox-research/Screen-Life-TV-in-demand>. 2013; Posted July 18, 2013.
4. Woods RL, Satgunam P. Television, computer and portable display device use by people with central vision impairment. *Ophthalmic Physiol Opt*. 2011;31:258–274.
5. Costela FM, Saunders DR, Rose DJ, Kajtezovic S, Reeves S, Woods RL. People with central vision loss have difficulty when watching videos. *Invest Ophthalmol Visual Sci*. 2019;60:358–364.
6. Fullerton M, Peli E. Digital enhancement of television signals for people with visual impairments: evaluation of a consumer product. *J Soc Inf Display*. 2008;16:493–500.
7. Fullerton M, Woods RL, Vera-Diaz FA, Peli E. Measuring perceived video quality of MPEG enhancement by people with impaired vision. *J Opt Soc Am (A)*. 2007;24:B174–187.
8. Kim J, Vora A, Peli E. MPEG-based image enhancement for the visually impaired. *Optical Eng*. 2004;43:1318–1329.
9. Peli E. Recognition performance and perceived quality of video enhanced for the visually impaired. *Ophthalmic Physiol Opt*. 2005;25:543–550.
10. Al-Atabany WI, Memon MA, Downes SM, Degeenaar PA. Designing and testing scene enhancement algorithms for patients with retina degenerative disorders. *Biomed Eng Online*. 2010;9:27.
11. Peli E, Kim J, Yitzhaky Y, Goldstein RB, Woods RL. Wideband enhancement of television images for people with visual impairments. *J Opt Soc Am (A)*. 2004;21:937–950.
12. Alfano PL, Michel GF. Restricting the field of view: perceptual and performance effects. *Percept Motor Skills*. 1990;70:35–45.
13. Turano K, Schuchard RA. Space perception in observers with visual field loss. *Clin Vision Sci*. 1991;6:289–299.
14. Pelli DG. The visual requirements of mobility. In: Woo GC (Ed.), *Low Vision Principles and Applications*. New York: Springer-Verlag. 1986:134–146.
15. Hassan SE, Hicks JC, Lei H, Turano KA. What is the minimum field of view required for efficient navigation? *Vision Res*. 2007;47:2115–2123.
16. Lovie-Kitchin JE, Soong GP, Hassan SE, Woods RL. Visual field size criteria for mobility rehabilitation referral. *Optom Vision Sci*. 2010;87:E948–957.
17. Marron JA, Bailey IL. Visual factors and orientation-mobility performance. *Am J Optom Physiol Opt*. 1982;59:413–426.
18. Udagawa S, Ohkubo S, Iwase A, et al. The effect of concentric constriction of the visual field to 10 and 15 degrees on simulated motor vehicle accidents. *PLoS One*. 2018;13:e0193767.
19. Wood JM, Troutbeck R. Effect of restriction of the binocular visual field on driving performance. *Ophthalmic Physiol Opt*. 1992;12:291–298.
20. Bowers A, Peli E, Elgin J, McGwin G, Jr., Owsley C. On-road driving with moderate visual field loss. *Optom Vision Sci*. 2005;82:657–667.
21. Coeckelbergh TR, Brouwer WH, Cornelissen FW, Van Wolffelaar P, Kooijman AC. The effect of visual field defects on driving performance: a driving simulator study. *Arch Ophthalmol*. 2002;120:1509–1516.
22. Rosenholtz R. Capabilities and limitations of peripheral vision. *Ann Rev Vis Sci*. 2016;2:437–457.
23. Brady TF, Shafer-Skelton A, Alvarez GA. Global ensemble texture representations are critical to

- rapid scene perception. *J Exp Psychol Hum Percept Perform.* 2017;43:1160.
24. Ehinger KA, Rosenholtz R. A general account of peripheral encoding also predicts scene perception performance. *J Vision.* 2016;16:13.
 25. Ullman S, Assif L, Fetaya E, Harari D. Atoms of recognition in human and computer vision. *Proc Natl Acad Sci USA.* 2016;113:2744–2749.
 26. Dorr M, Martinetz T, Gegenfurtner KR, Barth E. Variability of eye movements when viewing dynamic natural scenes. *J Vision.* 2010;10:28.
 27. Goldstein RB, Woods RL, Peli E. Where people look when watching movies: do all viewers look at the same place? *Comput Biol Med.* 2007;37:957–964.
 28. Saunders DR, Bex PJ, Rose DJ, Woods RL. Measuring information acquisition from sensory input using automated scoring of natural-language descriptions. *PLoS One.* 2014;9:e93251.
 29. Saunders DR, Bex PJ, Woods RL. Crowdsourcing a normative natural language dataset: a comparison of Amazon Mechanical Turk and in-lab data collection. *J Med Internet Res.* 2013;15:e100.
 30. Brainard DH. The Psychophysics Toolbox. *Spat Vision.* 1997;10:433–436.
 31. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vision.* 1997;10:437–442.
 32. Costela FM, Woods RL. When watching video, many saccades are curved and deviate from a velocity profile model. *Front Neurosci.* 2018;12.
 33. Janssen DP. Twice random, once mixed: applying mixed models to simultaneously analyze random effects of language and participants. *Behav Res Meth Instr Comput.* 2012;44:232–247.
 34. Reeves S, Williams V, Costela FM, et al. Narrative video scene description task discriminates between levels of cognitive impairment in Alzheimer's disease. *Neuropsychology.* 2020;34:437.
 35. Tseng PH, Carmi R, Cameron IG, Munoz DP, Itti L. Quantifying center bias of observers in free viewing of dynamic natural scenes. *J Vision.* 2009;9:4.
 36. Leat SJ, Rumney NJ. The experience of a university-based low vision clinic. *Ophthalmic Physiol Opt.* 1990;10:8–15.
 37. Deemer AD, Swenor BK, Fujiwara K, et al. Preliminary evaluation of two digital image processing strategies for head-mounted magnification for low vision patients. *Transl Vis Sci Technol.* 2019;8:23.
 38. Dickinson CM, Fotinakis V. The limitations imposed on reading by low vision aids. *Optom Vis Sci.* 2000;77:364–372.
 39. Lovie-Kitchin JE, Woo GC. Effect of magnification and field of view on reading speed using a CCTV. *Ophthalmic Physiol Opt.* 1988;8:139–145.
 40. Mohammed Z, Dickinson CM. The interrelationship between magnification, field of view and contrast reserve: the effect on reading performance. *Ophthalmic Physiol Opt.* 2000;20:464–472.
 41. Aguilar C, Castet E. Evaluation of a gaze-controlled vision enhancement system for reading in visually impaired people. *PLoS One.* 2017;12:e0174910.
 42. Calabrèse A, Aguilar C, Faure G, Matonti F, Hof-fart L, Castet E. A vision enhancement system to improve face recognition with central vision loss. *Optom Vision Sci.* 2018;95:738–746.