Evaluating Visual Information Provided by Audio Description

E. Peli, E.M. Fine, A.T. Labianca

Abstract: Short segments of two TV programs without audio description (AD) were presented to 25 subjects with low vision and 24 subjects with normal vision, and 29 additional subjects heard only the standard audio portions. The subjects then answered questions based on the visual information contained in the AD of the programs. The subjects with normal vision performed the best, followed by those with low vision and those who heard only the audio portion; all performed at better than chance levels. The results indicate that although AD may provide information on visual details to visually impaired audiences, some of the information in the AD may be obtained from the standard audio portion.

Television (TV) and other video outlets are an important means of obtaining information and sharing in the culture of a society. For people with visual impairments, the visual aspects of these media are impoverished or absent. Nevertheless, most visually impaired people watch TV with their families and use other video sources, such as videotaped movies and computer video displays (Berkowitz, Hiatt, deToledo, Shapiro, Lurie, 1979; Josephson, 1968). In fact, visually impaired people watch as much or more TV as do sighted people (Cronin & King, 1990).

Audio description (AD) provides descriptions of the visual elements of TV programs without interfering with the programs' audio (Cronin & King, 1990). Descriptions of details concerning such aspects of a scene as clothing styles, body language, colors, and landscapes are inserted during pauses in the dialogue or narration of normal programs. The AD is typically broadcast via the separate audio program (SAP) channel that is a standard feature of most stereo TVs and videocassette recorders (VCRs). ADs of programs broadcast on the SAP are available free of charge and do not interfere with other viewers' enjoyment of the programs. AD is also available with VCRs, for which narration is simply added to the videotapes, which are played on standard equipment. Although the reviews by visually impaired audiences have been positive, AD was not formally evaluated with visually impaired people until the study presented here.

Rabbitt and Carmichael (1993) examined the comprehension by older adults with normal vision of short segments of programs, as well as of an entire 30-minute program, presented to half the subjects

This research was supported by NIH Grant EY05957 to E. Peli. The authors wish to thank Julie Irish for her help in developing the questions used in this study; Barry Cronin and Gerry Field, of WGBH, Boston, for providing the program videos, software, and valuable technical help; and Laurie A. Everett, of WGBH's Descriptive Video Service, for her comments on an earlier draft of the article.

with AUDETEL, a British AD service, and to half without AUDETEL. They tested the subjects' comprehension of each segment by asking them to identify characters and locations that appeared in the segment and to elaborate on their descriptions. They also measured the subjects' overall comprehension of the program using 20 true-false questions.

Rabbitt and Carmichael found that comprehension of the segments (but not overall comprehension) was greater for the group that was exposed to the AD and surmised that AD helped cue the subjects to features of the program that were the focus of the open-ended questions. The primary impetus for their study was to determine whether the AUDETEL descriptions were disruptive to elderly viewers. On the basis of their data, they concluded that AUDETEL descriptions did not detract from the enjoyment or interfere with the comprehension of the program by elderly viewers with normal vision.

The current study was designed to evaluate only whether and to what extent the visual information supplemented by AD is not available to persons with visual impairments. (AD occasionally includes information that is not strictly visual, such as period and temporal references, relationships, and poetic expressions, but these types of information were not included.) Although it seems evident that any such description would be useful for blind audiences, the authors were interested in the level of transfer of information from the visual content of the programs only to audiences with low vision.

To evaluate the potential for transfer of visual information from AD to visually impaired audiences, the authors asked volunteers with low vision and those with normal sight to answer questions developed on the basis of visual information provided by Descriptive Video Service (DVS), an AD service developed by WGBH, the public television station in Boston. Two different programs broadcast with DVS (one from the *Mystery*! series and one from the *Nature* series) were chosen because the important visual elements they contained were different and spanned the type of programming generally aired with AD. Observers were presented only with the video and standard audio portions of the films...They then answered multi-choice questions on the content of the audio descriptions without actually hearing them. It was hypothesized that if the descriptions contain visual information available to normally sighted viewers but not to those with low vision, then observers with normal vision would answer more of the questions correctly.

To verify that the questions could not be answered on the basis of the standard audio of the program or common knowledge, the authors asked an additional group of volunteers the same questions after hearing only the standard audio portion of the programs; these volunteers neither saw the video nor heard the AD. Although this group was included primarily to verify that visual information was needed to answer the questions, the data they generated can also be used to estimate the minimum level of performance that could be expected from blind audiences engaged in the same task.

Methods

Subjects

Adults aged 55 or older who were recruited to participate were separated into two groups on the basis of the acuity in their better eye. Normal vision (NV) was defined as acuity of 20/40 or better in the better eye and low vision (LV) as acuity between 20/100 and 20/800 in their better eye. Adults who did not meet the criterion for either group were excluded.

A total of 24 NV subjects (9 who viewed only *Nature*; 9, only *Mystery!*; and 6, both) and 25 LV subjects (10 who viewed only *Nature*; 10, only *Mystery!*; and 5, both) participated. There was no difference in mean age between the vision groups (NV = 71, SEM = 1.63; LV = 75, SEM = 2.10; $t(45) = -1.51$, $p = .14$). The median acuities for the NV and LV groups were 20/30 and 20/200, respectively.

Twenty-nine additional subjects (2 who heard only *Nature*; 7, only *Mystery!*; and 20, both) were recruited to participate in the audio-only (AO) condition. The age of this sample [mean age = 51, SEM = 4.15] was not restricted, and the authors continued to collect data from the AO group beyond the number of subjects in the NV and LV groups for another study. Since both the audio portions and the ADs of programs rely on hearing, and many visually impaired people are elderly and may have some hearing loss, the authors explored whether age might affect performance on this task. There was no statistical difference in performance between the 55-or-older group and the younger group [$t(47) < 1.0$, n.s.], and so decided to maintain the larger sample size for the analyses.

The AO group neither saw the video nor heard the AD, so their responses were based only on the standard audio of the program. Since the purpose of this condition was to verify that questions could not be answered on the basis of the standard audio alone. these subjects' acuity was not measured. It should be noted that the AO condition is not a simulation of a truly blind audience, which may have adapted to long-term vision loss by gathering information through other sensory inputs (primarily auditory). Thus, it is likely that the AO group would have been outperformed by a group of blind observers who are presumably more skillful at gaining information from nonvisual sources.

Apparatus

An Apple Macintosh SE was used to run a HyperCard® stack, de**Script**or version 1.2.1 (DVS, Boston). Modifications to this stack allowed the authors both to control the presentation of the video and to record subjects' responses. The computer was interfaced with a Sony U-matic/SP VCR via two Videomedia Serial I/O units that synchronized the time code between the computer and the VCR.

DVS de**Script**or stacks were originally developed to add AD to TV programs. Using this system, a describer can control the VCR to navigate precisely through and display particular scenes. In addition, a description can be added by advancing to the appropriate scene and adding the narration to a script field designated for that scene. The authors added question-and-response fields in lieu of the narration script field that DVS created. They also modified the software to export response data automatically from the response fields to spreadsheet format. These modifications reduced data-entry errors, as well as any risk that the experimenter would ask questions or record responses that did not correspond to the correct scene or question.

The video segments were displayed on a 27-inch Sony Trinitron color TV monitor. Acuity was tested monocularly using a

Mentor B-VAT II, a system in which sans serif letters are randomly chosen from a limited set and can be displayed individually on a 12-inch monitor.

Materials and design

Since DVS is available to TV audiences across the country both by broadcast media and by the production of videotapes with an AD track, episodes of two TV programs that had been broadcast with DVS descriptive narration were chosen: "Seasons in the Sea" from the *Nature* series and "Poirot: The Theft of the Royal Ruby" from the *Mystery!* series. A continuous series of segments of each episode (approximately 10 minutes in length) was presented with only their normal audio (without the DVS narrations). The addition of AD is normally constrained by the breaks in the standard audio of the program. AD can occur simultaneously with, before, or after the visual element described. When possible, the episode was paused to ask questions about a segment at the same point that DVS parsed it for inserting narration. When the visual portion did not coincide with the DVS script, a segment of the video that contained the portion corresponding to that DVS narration was presented. Portions of the video that did not directly correspond to the questions were also included to maintain the continuity of the episode. The continuous 10 minutes of film was parsed into 21 and 17 segments for *Nature* and *Mystery!*, respectively.

DVS descriptions were used as a basis for choosing the visual details that are assumed to be unavailable to persons with visual impairments. Although questions were not developed from the entire DVS script, the 10-minute segments chosen from each episode were representative of the descriptions used throughout the program. Two-alternative forced-choice questions (43 for the *Nature* episode and 59 for the *Mystery!* episode) were generated on the basis of the DVS descriptions of the corresponding scenes (see Box 1 for examples). These questions were designed to test whether a visual detail described by DVS was seen or not seen. After a third party developed the questions, the first and third authors examined the DVS script and the video portions of the programs to determine whether the questions did, in fact, ask about specific features both contained in the segment and described by DVS. When they disagreed about the validity of a question, they decided either to eliminate it or to modify it. Because the number of visual details described in the DVS narration for each scene differed, the number of questions per segment varied. There was an average of 2.0 questions per segment for the *Nature* episode and 3.5 questions per segment for the *Mystery!* episode.

Nature	*Mystery!*
From the side, the jellyfish's red cap surges gracefully forward, towing ruffled tentacles with rose-tinted edges.	A smooth-faced, gray-haired woman. She glances at the door. She paces hesitanlty across the drawing room.
1. The jellyfish's arms are . . . a. smooth b. ruffled	1. The woman has . . . a. many wrinkles b. a smooth face
2. The arms have . . . a. greenish blue edges b. reddish pink edges	2. The woman has . . . a. gray hair b. black hair

Box 1. Examples of AD and corresponding questions.

Acuity testing

Acuity was measured monocularly with the same refractive correction used for viewing the TV programs and with the room lights off, using a Mentor B-VAT II. For each eye, subjects were asked to name letters, beginning at a letter size corresponding to 20/600, that appeared individually on the screen. They stood 10 feet from the monitor, which was adjusted for testing from 20 feet. This combination allowed verification of acuity from 20/30 to 20/600. The target size was decreased until fewer than four of five letters were correctly identified. The size was then increased, one step at a time, until four of the five letters were named correctly. This target size was recorded as the acuity for the eye tested. Subjects who were unable to identify letters at the 20/600 size were moved closer to the screen (5 feet), thus increasing the relative character size to 20/1200. If their acuity was better than 20/30, the B-VAT II was adjusted for testing from 10 feet, which allowed verification of acuity down to 20/15.

Procedure

Presentation of programs

After the acuity testing, each subject was told that he or she would watch two short portions of TV programs that were parsed into several segments and answer multiple-choice questions based on each segment. The presentation order of the two programs was counterbalanced. The volume was adjusted to accommodate each subject. Each subject was seated approximately 6 feet from the TV monitor and wore his or her usual refraction for distance viewing. Although some low vision viewers may normally sit closer to their TVs, the authors chose to standardize the viewing distance.

With the room lights on, the video segments were presented without the AD. After presenting each segment with only the regular program audio, the experimenter read the questions and possible answers to the questions that corresponded to the segment just presented and recorded the subject's verbal responses. When multiple questions were derived from a single segment, each question and its possible answers were presented, and the subject's response was recorded before the next question was read. The questions always followed the order of events in the DVS narration. When unsure, subjects were asked to guess, and all questions were answered by all subjects. Several subjects completed only the first program because of time constraints. Approximately 45 minutes were required to administer each program. The same procedure was followed for the AO group except that the television was turned to face the wall.

Results and discussion

Since the purpose of the study was to determine the amount of visual information obtained from the video by persons with low vision, questions that could be answered on the basis of the standard audio of the program and questions whose correct answer could not be chosen based on normal viewing were eliminated from the analyses. A strict criterion was used for exclusion, based on the performance of both the AO and NV groups, to eliminate a total of five questions. Four questions (2 from *Mystery!* and 2 from *Nature*) were eliminated because they were answered correctly by 90 percent or more of the AO subjects, suggesting that they could be answered by listening to the standard audio alone or on the basis of common knowledge. The remaining question (from *Mystery!*) was eliminated because it was answered correctly by 10 percent or fewer of the NV subjects, suggesting that it could not be answered on the basis of normal viewing.

The proportion of questions answered correctly by each subject for each episode (see Table 1) was computed, and a 3 (vision group: NV, LV, or AO) by 2 (program: *Mystery!* or *Nature*) ANOVA was used to evaluate the data. Because not all the subjects in each condition saw both programs, the data from each viewing of each program were treated as independent in the analyses. Although this is an unconventional method of analysis, it results in an increase in the estimates of variability within each vision group and thereby decreases the likelihood of finding differences among these groups.

The results of the ANOVA indicated a main effect of vision group [$F(2,103) = 130.59$, $p = .0001$] and a main effect of program [$F(1,103) = 6.77$, $p = .0107$], but no interaction between these variables [$F(2,103) = 1.84$, $p = .1639$]. Bonferroni/Dunn's post-hoc analyses showed that the NV group performed better than both the LV and AO groups and that the LV group performed better than the AO group for both programs ($p = < .0005$ for all analyses). This finding indicates that at least some of the visual information contained in the AD descriptions can be attained by low vision viewers and that some benefit can be derived from these descriptions. Had all the visual information contained in the AD been available to the low vision viewers, it is unlikely that their performance would have been reduced relative to the NV group.

The AO group was included primarily to assess the validity of the questions, but their data can be used to estimate the minimum level of performance that could be expected from blind audiences engaged in the same task. It is evident from the preceding analysis that when only the information contained in the audio portion of the programs was available to these subjects, their performance was reduced relative to both subjects with normal vision and with partial sight. However, their performance did not drop to chance levels (50%), even when questions with 90 percent or better performance were eliminated. Although the 4 percent and 5 percent differences from chance performance for the AO group answering questions from *Nature* and *Mystery!*, respectively, were small, they were statistically greater than chance ($t > 3.0$, $p < .006$ for both). Two conclusions can be drawn from this finding. First, the AD contains some visual information redundant to the standard audio or common knowledge, and second,

Table 1
Mean percentage of questions answered correctly by three groups of observers (NV, LV, and AO) in each episode presented (*Nature and Mystery!*).*

	Nature			*Mystery!*		
	NV	LV	AO	NV	LV	AO
n	15	15	23	15	15	26
Mean	79.4	69.8	54.1	87.3	73.0	55.0
SEM	1.7	2.5	1.3	1.2	3.0	1.5

* NV = normal vision, LV = low vision, and AO = audio only.

blind audiences would be able to acquire at least this much of the visual detail contained in the program without AD.

As was discussed earlier, the questions were developed on the basis of the visual details in the AD narration, portions of which described aspects of a scene that were evident without seeing the video portion. That almost 16 percent of the questions were answered correctly by 80 percent or more of the subjects in the AO condition suggests that some of the content of the AD could be identified from the audio portion alone or from common knowledge. For instance, the first sample question from *Nature*, shown in Box 1, can be answered by listening to the standard audio, which states: "A gigantic pelagic jellyfish . . . has extravagantly frilled oral arms trailing some 20 feet from its mouth." The corresponding description regarding the "ruffled tentacles" is redundant here.

Since persons with low vision frequently sit closer than 6 feet from the TV screen and use optical aids (in addition to normal refraction) to watch TV, it is likely that the overall performance of the LV group would have improved if each LV subject had been allowed to choose his or her own viewing distance. If it had, it would have reduced the difference between the NV and LV groups and would more strongly suggest that the AD provides information that persons with low vision can obtain elsewhere. The same argument can be made about the performance of the AO group relative to a blind audience. That is, a blind audience would be likely to outperform the AO group because they have adapted to the loss of sight and are more adept at gaining information from auditory inputs. Again, their performance would diminish the difference between groups and strengthen the conclusion that some portions of the AD are redundant.

Conclusion

This investigation has shown that without hearing AD, low vision observers and persons who were exposed only to the audio portion of the programs answered fewer questions correctly than did normally sighted viewers. This finding suggests that the information contained in AD narrations may provide benefits specifically related to the availability of visual details to both visually impaired and blind viewers. There are also indications, however, that some of the AD narrative was redundant with the audio portion of the programs. Although the inclusion of redundant and extra information is not inherently negative, the time may be better used to describe elements of programs that visually impaired viewers and blind audiences are unable to obtain elsewhere. However, if the extra information provided by AD is necessary for visually impaired and blind viewers to follow a story or enjoy a program, AD may be an ideal sensory substitute for viewers who cannot fully appreciate visual details.

Totally blind people represent a small minority of the visually impaired population. Most visually impaired people have partial sight, and the majority are elderly and have some degree of hearing loss in addition to their visual impairment (Rabbitt & Carmichael, 1993). Thus, many of the viewers for whom AD is designed may not be able to benefit from it. Furthermore, the availability of AD is likely to remain limited (currently, only about 10 PBS programs per week are broadcast with DVS), and its applicability is generally limited to off-line production, although live broadcasts, such as U.S. President Clinton's inauguration, which was broadcast with AD, are possible. In addition, the production of AD is time consuming and costly relative to other services (such as closed captioning) that are designed for audiences with disabilities, although not in relation to overall production costs. For example, the production and narration costs of adding DVS to a PBS documentary are about two-thirds of 1 percent of the total production costs, and the costs of adding AD to a videotaped movie are only about one-fiftieth of 1 percent (personal communication with L.A. Everett, DVS, July 24, 1995).

Despite some limitations, with today's technology, AD may still be the best solution for blind users. However, there may be alternatives for low vision users. For example, contrast enhancement has been shown to improve the recognition of faces by persons with low vision (Peli, Lee, Trempe, & Buzney, 1994) and to increase viewers' appreciation and identification of details of videotaped movies (Peli, Fine, & Pisano, 1994). Using the questions developed for this study, the authors will be able to evaluate systematically the efficacy of contrast enhancement for low vision users. If it is found that enhanced visual details are identified by low vision viewers with at least the same accuracy as with AD, then contrast enhancement would be a better alternative for these viewers because they can apply it themselves to any video output, including broadcast TV, VCR output, and computer-generated displays.

References

Berkowitz, M., Hiatt, L.G., deToledo, P., Shapiro, J., & Lurie, M. (1979). *Characteristics, activities and needs of people with limitation in reading print*. New York: American Foundation for the Blind.

Cronin, B.J. & King, S.R. (1990). The development of descriptive video service. *Journal of Visual Impairment & Blindness*, **80**, 503–506.

Josephson, E. (1968). *The spatial life of blind people* (Research Series 19). New York: American Foundation for the Blind.

Peli, E., Fine, E.M., & Pisano, K. (1994). 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., Lee, E., Trempe, C.L., & Buzney, S. (1994). 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.

Rabbitt, P.M.A. & Carmichael, A. (1993). Assessment of the effect of audio description on elderly people's comprehension and memory for television programmes. In *AUDETEL Project Final Report, Vol. 2: Optimization of an audio description service for visually handicapped and elderly people: Human and technical factors*. London: Independent Television Commission.

***Eli Peli, M.Sc., O.D.**, senior scientist, Schepens Eye Research Institute, and associate professor, Harvard Medical School, 20 Staniford Street, Boston, MA 02114; E-mail: <eli@vision.eri.harvard.edu>. **Elisabeth M. Fine, Ph.D.**, postdoctoral fellow, Wilmer Eye Institute, Johns Hopkins University, 550 N. Broadway, Baltimore, MD 21205; E-mail: <fine@lions.med.jhu.edu>. **Angela T. Labianca, M.A.**, research assistant, Schepens Eye Research Institute, Harvard Medical School, 20 Staniford Street, Boston, MA 02114; E-mail: <ladigreg@vision.eri.harvard.edu>.*

JVIB

Journal of Visual Impairment & Blindness

September-October 1996 Volume 90 Number 5

ARTICLES

DEPARTMENTS

0145-482X(199609/10)90:5;1-4

Contents

ARTICLES

DEPARTMENTS

The *Journal of Visual Impairment & Blindness* (ISSN: 0145-482X) is published bi-monthly by AFB Press, American Foundation for the Blind, 11 Penn Plaza, Suite 300, New York, NY 10001; individual subscriptions, $59 per year, institutional subscriptions, $87 per year. Periodicals postage paid at New York, New York and at additional post offices. POSTMASTER: Send address changes and subscription orders to *Journal of Visual Impairment & Blindness*, 49 Sheridan Ave. Albany, NY 12210.

AFB PRESS