Non-Speech Information in Captioned Video: A Consumer Opinion Study

with Guidelines for the Captioning Industry

By
Judith E. Harkins, Ph.D., Project Director
Ellie Korres, Research Associate
Beth Singer, M.S., CCC-A, Research Associate
and Barbara M. Virvan, M.S.W., Research Associate

Technology Assessment Program
Gallaudet Research Institute
Gallaudet University

based on a study funded by
The U.S. Department of Education
Office of Special Education Programs
Award #H026R20004
and
Gallaudet University

Persons interested in obtaining additional copies of this document or of the Gallaudet Research Institute newsletter *Research at Gallaudet* *may write to:*

Robert Clover Johnson, Research Editor
The Graduate School and Research
HMB S449, Gallaudet University
800 Florida Avenue, N.E.
Washington, D.C. 20002

Phone (V/TTY): (202) 651-5400
FAX: (202) 651-5295

Table of Contents

Introduction 1
- Summary of Methods used in this Study 2
- Summary Results 3
- Summary Table 1: Guidelines for Non-Speech Information 5
- Summary Table 2: Guidelines for Features 7

Data Collection Procedures 8
- Data Analysis 8

Results and Recommendations 9
- Speaker Identification 9
  - Off-screen narrator 9
  - Multiple speakers, on- and off-screen 10
- Sound Effects 18
- Music 20
  - Background music 20
  - Singing 20
- Audience Reaction 24
  - Reaction other than laughter 24
  - Laughter 24
- Manner of Speaking 27
  - Emotion 27
  - Whispering 27
  - Emphasis 28
  - Accents 28
- Explanation of a Pun 33
- Indication of a Title 33
- Quantity of Indicators 36

References 38

Appendix A: Questionnaire and Results of Awareness of Features

Appendix B: Demographic Characteristics of the Sample

Appendix C: Companies and Organizations that Responded to the Draft Guidelines

Appendix D: Advisors, Consultant, and Staff

Introduction

A floorboard creaks in the hallway of a darkened house. Eerie music begins to play. A woman screams from off camera. A siren wails in the distance.

In television programs and movies, not all information is conveyed through dialogue. In fact, as shown in the example above, a great deal of information can be imparted through sound effects, music, manner of speaking and other kinds of "non-speech information."

Non-speech information (NSI) is a term that describes aspects of the sound track, other than spoken words, that convey information about plot, humor, mood, or the meaning of a spoken passage.

Examples of NSI include:

identification of speaker (off-screen speakers and multiple on-screen speakers)
sound effects
music (singing, background music, etc.)
manner of speaking (whispering, emotion, word emphasis, etc.)
audience reaction (laughing, groaning, booing, etc.)
indication of title (books, films, newspapers, plays, etc.)
puns

Many companies in the captioning industry are aware of the role played by this sound-based information, and understand its importance for access by deaf and hard of hearing audiences. However, companies often vary in the ways they portray this information.

A number of tools are used to indicate NSI. Recent changes in decoder circuitry permit caption writers more features for indicating NSI, and with this greater latitude comes the potential for even greater inconsistency among companies. Examples of features that can be used to identify NSI are:

italics/slanted type
placement of the caption near the speaker or sound source
upper/lower case letters
chevrons (>>)
color
music icons
paint-on captions
underlining
quotation marks
explicit description

Guidelines for captioning non-speech information would make it easier for deaf and hard of hearing viewers to follow captions. Toward that end, the Technology Assessment Program at Gallaudet University studied deaf and hard of hearing people's preferences for captioning NSI. The results of the study were used to develop recommended style guidelines to the captioning industry. Draft guidelines were sent to captioning companies for comment, and that input was incorporated into the final guidelines. The purpose of these guidelines is to improve captioning of NSI.

While the vast majority of recommendations in this report are based on a study conducted by the Technology Assessment Program between December, 1992 and May, 1994 some of the recommendations are further bolstered by findings of Cynthia King, Ph.D., and Carol LaSasso, Ph.D., of Gallaudet University's School of Education and Human Services. Both studies were funded by the U.S. Department of Education, Office of Special Education Programs, Captioning and Adaptation Branch. The results of the study apply to Line 21 captioning.

Summary of Methods used in this Study

- An **advisory committee** composed of consumers and caption industry representatives advised the research staff on all aspects of the study, and reviewed the findings and draft recommendations before they were circulated to the industry for comment.

- **Thirty-three caption writers were interviewed** for input as to the challenges they face in representing NSI.

- **Thirty-eight hours of video were analyzed in detail** to determine current practice in identifying NSI.

- **Nineteen video clips from television were selected** for the data collection videotape, providing 19 different examples of NSI. Emphasis was on NSI identified by the advisory committee and industry as being most important, such as speaker identification.

- For each of the 19 examples, **two or three different ways of captioning to indicate NSI were selected**. New uses of captions, incorporating features made possible by updated caption circuitry, were included as well as conventional features such as italics. In all, **55 uses of caption features were included**. The order of choices was counterbalanced, to eliminate order effects in preference selection. Note: None of the clips was a real-time captioned segment.

- **Deaf (n = 106) and hard of hearing (n = 83) consumers viewed the videotape and indicated their preference** from the choices presented. Respondents were

required to make a choice, and could write comments in addition. If any of the choices were unacceptable, consumers were instructed to mark those choices with an X.

- Prior to viewing the tape, consumers completed a brief **demographic questionnaire**, and also answered a questionnaire to determine their **awareness and recognition of the meaning of commonly used caption features**, such as italics and upper/lower case letters.

- **Results were analyzed and recommendations drafted.** These were reviewed by the advisory committee and were circulated in the fall of 1994 to the industry and caption funding agencies for comment.

Summary Results

- The **advisory committee** recommended that identification of speaker be the highest priority type of non-speech information studied. The group recommended testing features that are already in use, new features, and even features that were not believed by the investigators to be desirable. Awareness questions were added to the study at the suggestion of the advisory committee.

- **Caption writers** welcomed guidelines that are based on consumer-based data. Only about half of those interviewed used a stylebook or other written guidelines. Caption writers identified numerous problem areas in captioning. Identification of speaker was the NSI that generated most of the challenges for caption writers.

- The median age of consumer-respondents in this study was 40. Respondents ranged in age from 14 to 84. Sixty-one percent were female, and 39% male. Nineteen percent were members of minority groups. Half had college degrees and half did not. Hearing loss was reported as severe or profound for 70% of respondents. Age at onset of hearing loss was over 20 years of age for 26% of the sample; 52% became deaf at or before three years of age. Ninety-three percent had closed caption decoders in the home at the time of the study.

- **Consumers could not name the functions of commonly used caption features,** although such features as italics, upper/lower case, and double chevrons (>>) are very widely used to indicate NSI. Only a minority of consumers gave partially or completely correct responses to an evaluation of awareness of features. This situation may be due to non-standard industry practice and/or the fact that features used are somewhat difficult to interpret if one does not have access to the sound track. This result indicates that caption companies have little to lose by changing the specific features they use in favor of guidelines presented here.

- **Explicit speaker identification** (by name) was tested six times in the study. Two thirds of respondents chose it more than half of the time, whereas only 3% never preferred it to other features.

- **Description** was tested seven times. The vast majority of consumers picked description at least half of the time it was offered: 84% chose description four or more times.

- **Unusual uses of caption features** (not conventional in the industry at this time) were presented 11 times. Here the preferences were fairly normally distributed; that is, there was no tendency for people automatically to reject new ways of captioning. Specifically, two-thirds of respondents preferred an unusual feature five to seven times out of the 11 examples.

- **Conventional captioning styles**--those examples that were captioned as they actually appeared on the air--were presented 11 times. The distribution was somewhat more skewed. Three-quarters of respondents chose the default version one to four times out of 11. This trend was in part due to the fact that the version that aired did not indicate NSI in several cases, and consumers rejected the lack of indication.

- **Color** was tested five times. Only 19% chose it more than half the time it was offered. Twice as many--38%--never chose it, and 29% volunteered that it was unacceptable at least once.

Table 1 summarizes guidelines resulting from the project. Table 2 summarizes recommendations regarding features. **Detailed descriptions of the findings leading to these recommendations appear after Tables 1 and 2.**

Summary Table 1
Guidelines for Types of Non-Speech Information

General guideline	**If a descriptive caption or feature would in any way clarify or enhance the viewer's awareness of the audio, it should be indicated. Consumers prefer that more of such information be included than is often done in current practice.**
Background music	**Background music should be indicated, especially if it contributes to the plot or mood of the video. A description of the background music should be given wherever possible.**
Sound effects	**Where feasible, a combination of description and onomatopoeia should be used to indicate sound effects. If space or other limitations do not permit the two to be used together, descriptors should be used. Onomatopoeia should not be used alone. A descriptor is particularly important if the source of the sound effect is not obvious from the video.**
Singing	**Continue the practice of using the musical-note icon surrounding the caption. All-caps and upper/lower-case type are equally acceptable for the caption portion.**
Multiple speakers on screen	**Where multiple speakers appear on the screen, placement should be used to distinguish among them. Explicit identification should be used in combination with placement if dialogue is fast, if faces are obscured, if characters are moving, or if other circumstances could confuse the viewer. If the character cannot be identified by name, then a descriptor should be provided.** **An acceptable format for explicit identification is the character's name or descriptor in upper/lower case, surrounded by parentheses, above the caption and left justified with the caption. Minor variations of this format are probably uncontroversial.**
Narrators	**Explicitly identify off-screen narrators, rather than using features, such as italics or color, that require the viewer to interpret the feature/code while reading captions.**
Whispered speech	**Whispered lines should be identified as such and combined with upper/lower case captions.**

Emphasis of a word or phrase within a caption	Indicate the emphasized word(s) within a caption with italics.
Titles	Use quotation marks when indicating the title of a book, movie, etc.
Audience reaction	Audience reaction should be captioned. This is particularly important where the reaction itself becomes part of the plot or comedy. Audience laughter should also be described. (It is of course possible that repeating the descriptor every time the audience laughs, over the length of an entire sitcom episode, would become annoying. This length of exposure was not tested. Therefore, discretion is advised; but audience laughter should be indicated much more often than is now the industry's practice.)
Conveying emotion	Where strong emotion is conveyed, the emotion should be described with the caption. This feature should be used especially where the strong emotion is not entirely obvious in the facial expression and actions of the speaker. Caption writers may be concerned that this feature could be overused. However, based on consumers' reaction, caption writers should use this feature more than is current practice.
Accents	Indicate foreign or regional accent with a one-time description at the beginning of the character's lines. (Note: This issue was tested only with a fictional character, and probably should not be generalized to other speakers.)
Puns	Puns should be briefly explained when feasible.

Summary Table 2
Guidelines for Features

General Guideline	Consumers have indicated a preference for explicit description or identification over features that assume understanding on the part of the viewer. Examples of such features, requiring interpretation by the viewer, include: use of italics for the entire caption, color, and upper and lower case type without explanation.
Color	Color was not the preferred method of indication in this study, although it was tested in five different circumstances. Color also tested poorly against placement and speaker identification in an earlier study by King and LaSasso (1993). Color is judged unacceptable by more viewers than are many other features. Note that color in real-time captioning (where other options may be problematic) was not tested. (Color in a digital video environment is being studied further by King and LaSasso in 1994-1996).
Flashing	Flashing captions were not preferred in the two applications tested in this study, and were unacceptable to an appreciable minority of respondents. Further study may be warranted of whether or how to use this feature.
Paint-on	Paint-on captions were tested in only one context, and they were not preferred. Further study may be warranted of whether or how to use this feature.
Italics	Italics were less desirable than explicit definition in several contexts. Italics are widely used and should be used less frequently, as their intent is often lost on viewers.
Underline	Underlining was the last choice of respondents in the two applications tested. Further study may be warranted of whether or how to use this feature.
Quotation Marks	Quotation marks were preferred (contrasted with italics and underlining) for indicating a title.

Data Collection Procedures

Data were collected in 28 sessions in the fall of 1993, in seven states on the east coast of the U.S. Respondents were recruited through community groups, schools, and agencies.

Each respondent completed a questionnaire on awareness of caption features. The results of this questionnaire are included in Appendix A. Each respondent also completed a questionnaire on basic demographic information. The results of this questionnaire are included in Appendix B.

Respondents then watched 19 clips of video, each of which was captioned two or three different ways with regard to NSI. Clips ranged in length from 12 to 54 seconds. The order of choices was counterbalanced, to eliminate order effects. The response forms reviewed the feature used in each version of the clip, so that respondents did not have to rely on memory to make a selection.

In deaf groups, the instructions were printed and presented in American Sign Language. In hard of hearing groups, loop amplification was used during the spoken explanation, and print instructions were also provided. The instructions follow:

- You will see scenes from television shows.
- Each scene will be shown three times. (A few will be shown only twice.)
- The scene will be captioned a different way each time.
- You will have an answer sheet for each scene.
- The answer sheet has three choices, one for each different way of captioning.
- Please circle the one you prefer; you must choose one.
- You can cross out any that you think are unacceptable.
- You may ask questions at any time.
- Please do not discuss your choices with other people in the room. We want each person's individual opinion, not group opinions.

Data Analysis

Data were entered into a database and analyzed using Statistical Package for the Social Sciences. Frequencies and relative frequencies were obtained for demographic information, awareness questions, and preference questions. Chi-square analysis was conducted to test the hypothesis that the obtained distribution of preferences was different from the statistically expected distribution of preferences -- that is, to determine whether the pattern of preferences was not likely to be by chance. Data on preferences were also aggregated across clips to determine the preference or rejection of certain types of features, such as color, description, or new uses of captions.

Results and Recommendations

Speaker Identification

Speaker identification is one of the most important categories of non-speech information, because problems in identifying the speaker are frequent and can cause confusion.

A study by King and LaSasso found support for placing captions near the speaker (King and LaSasso, 1993). In this study, therefore, placement was accepted as a desirable feature, and was kept constant--that is, was used in every version tested.

Off-screen narrator

Off-screen narration was studied in two clips. In one clip tested, there were two off-screen narrators. The second narrator was identified as follows: (1) by explicit identification, (2) by color, and (3) by italicized upper/lower case text. Explicit identification was the rather strong favorite (67%) and was unacceptable to no one. In this clip, color was preferred by 21% and the use of italicized upper/lower text was preferred by only 12%. ($\chi^2 = 96.793$, $df = 2$, $p < .001$)

In another clip, the speech was produced by a narrator and a character who sometimes spoke off-screen. Features tested were: (1) explicit identification, (2) color, and (3) italicized caps. Explicit identification was again favored (65%) and found unacceptable by only one person; color was preferred by 19% and capital text with italics by 18%. ($\chi^2 = 83.079$, $df = 2$, $p < .001$)

In these two examples, color was unacceptable about as often as it was preferred.

Recommendation: Explicitly identify off-screen narrators, rather than using features, such as italics or color, that require the viewer to interpret the feature/code while reading captions.

Example:

> (Female narrator)
> THIS IS A GREAT DAY FOR THIS TEAM.

Multiple speakers, on- and off-screen

Features for multiple speakers were tested four times. All features were tested in combination with placement of captions near the speakers.

In one clip, a conversation among three characters included close-ups of the main character's face while the others were speaking. Speakers were identified as follows: (1) by explicit identification, (2) by color, and (3) by the use of double chevrons (>>) before each new speaker--an industry convention in real-time captioning--with placement. Explicit identification was preferred (56%) by the majority of respondents. Color came in second (31%) but was unacceptable to 16%. Chevrons were preferred by only 13%. ($\chi^2 = 51.175$, $df = 2$, $p < .001$)

In another clip, all speakers were on screen, but their faces were obscured, making it difficult to identify the speaker visually. Features tested were (1) explicit identification (with placement), (2) color (with placement), and (3) placement without additional features. Explicit identification was preferred by 64% of the respondents, followed by color (23%) and placement alone (13%). ($\chi^2 = 80.222$, $df = 2$, $p < .001$)

In a third clip, several characters spoke from off-screen. Three methods of explicit identification were tested: (1) using parentheses around the speaker's name; (2) speaker's name followed by a colon; and (3) brackets around the speaker's name. Preferences were not strong in this case: Parentheses were preferred by more respondents (41%) than were use of a colon (30%) or brackets (29%). Brackets were somewhat more unacceptable (6%) than were parenthesis (2%) or colon (3%). Because the differences were not statistically significant ($\chi^2 = 4.698$, $df = 2$, $p < .10$) the results are not used for guidelines.

In the fourth clip, actors re-created a scene based on a recording of an emergency telephone call. Both speakers were off-screen at all times. Features contrasted were (1) speaker identification in combination with placement, all caps; (2) speaker identification combined with all caps and upper/lower case; and (3) placement combined with all caps and upper/lower case but no speaker identification. Again, explicit identification of the speaker was the first choice, but differences in preference were statistically non-significant. [Speaker identification with placement and all capitals was preferred (41%) over upper/lower case text without placement (30%) and over the use of font (caps/upper-lower) (29%) without speaker identification.] Because the differences in preference for this clip were not statistically significant ($\chi^2 = 5.429$, $df = 2$, $p < .10$) the results are not used for guidelines.

Recommendation: Where multiple speakers appear on the screen, placement should be used to distinguish among them. Explicit identification should be used, particularly if dialogue is fast, if faces are obscured, if characters are moving, or if other circumstances could confuse the viewer. If the character cannot be identified by name, then a descriptor should be provided.

Recommendation: An acceptable format for explicit identification is the character's name or descriptor in upper/lower case, surrounded by parentheses, above the caption and left justified with the caption. Other formats are probably uncontroversial.

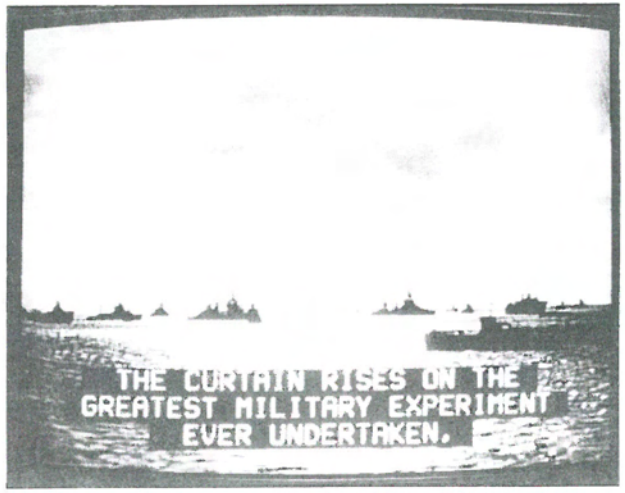
Examples:

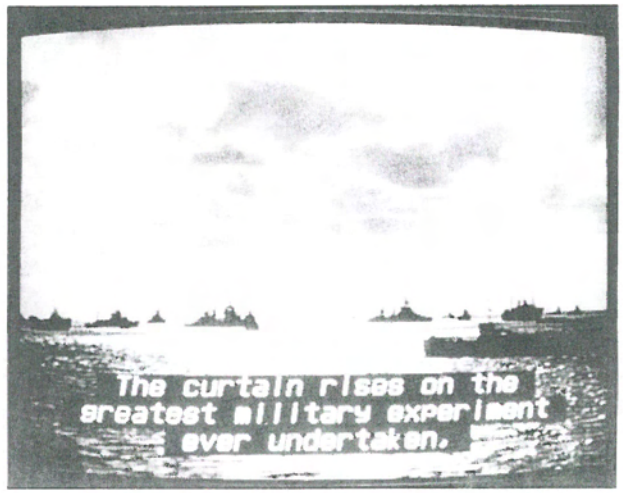
(Commander)
IN OUR OWN POLICE SHUTTLE.

(Tess)
PLEASE DON'T MAKE ME DO THAT.

SPEAKER IDENTIFICATION
OFF-SCREEN NARRATOR

In this clip, there are two off-screen narrators.





Features used to indicate speakers:

Speaker ID
Narrator:

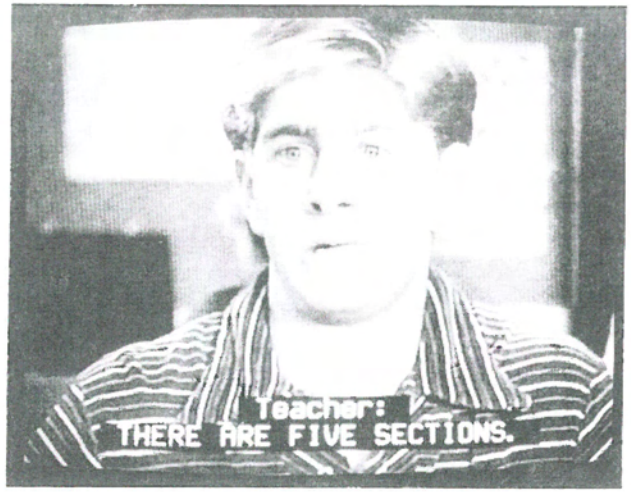
Color
Narrator in white, second speaker in yellow

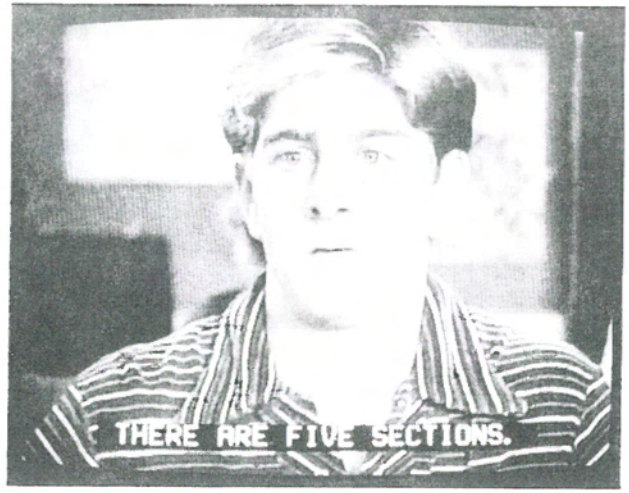
Capital letters and italics
CAPITAL LETTERS/*Slanted letters*

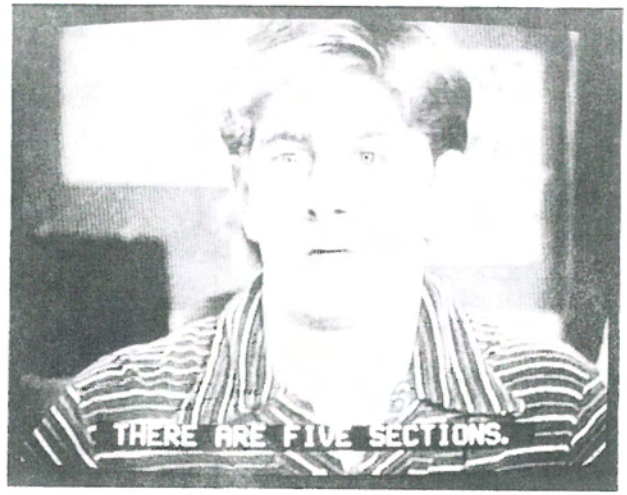
Feature	**Preferred**		**Unacceptable**	
	N	%	N	%
Speaker ID	126	67	0	0
Color	40	21	36	19
Caps & italics	23	12	10	5
Totals	189	100	46	24

SPEAKER IDENTIFICATION
OFF-SCREEN NARRATOR

In this clip, there are two speakers. One is a narrator. The other speaker is not always visible.







Features used to indicate speakers:

Speaker identification
Adult Kevin:

Capital letters and italics
SLANTED LETTERS/CAPITAL LETTERS

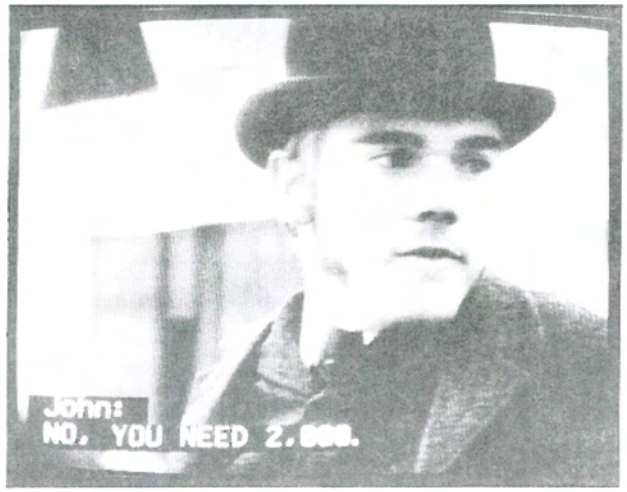
Color
Narrator in white, second speaker in green

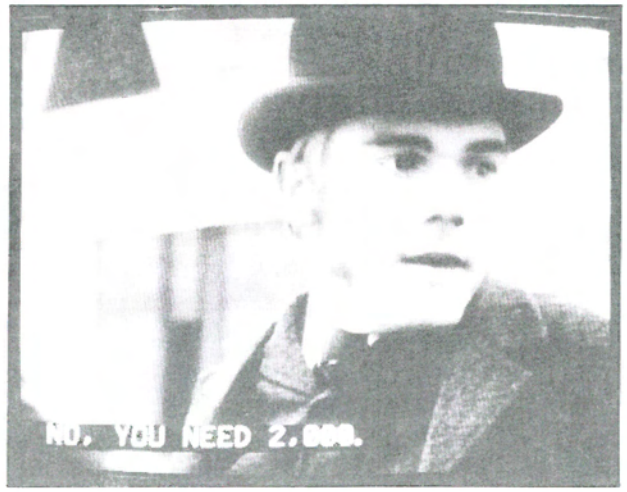
Feature	Preferred		Unacceptable	
	N	%	N	%
Speaker ID	122	65	1	1
Caps & italics	31	18	15	8
Color	36	19	36	19
Totals	189	101*	52	28

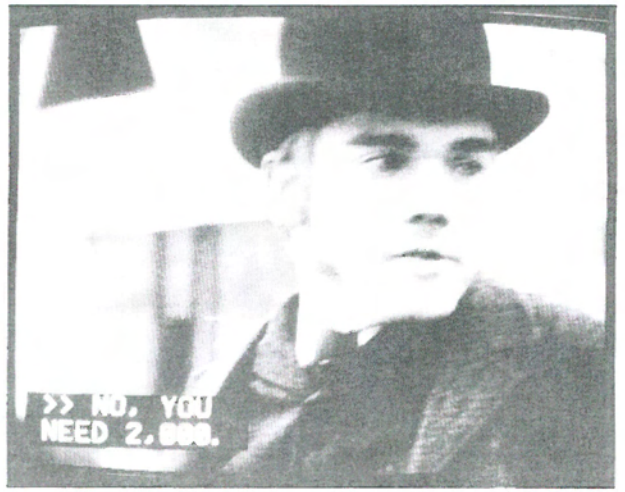
*Discrepancies of 1% are a result of standard rounding error.

SPEAKER IDENTIFICATION
MULTIPLE SPEAKERS

In this clip, there are multiple speakers, including more than one off-screen speaker.







Features used to indicate speaker:

Speaker identification
John:

Color
White, green and yellow used for the three speakers

Double chevrons
>>BEFORE EACH PERSON'S LINES

Feature	Preferred		Unacceptable	
	N	%	N	%
Speaker ID	105	56	3	2
Color	59	31	30	16
Double chevrons	25	13	20	11
Totals	189	100	53	29

SPEAKER IDENTIFICATION
MULTIPLE SPEAKERS

In this clip, the speakers are on-screen but hard to identify.

Features used to indicate speaker:

Placement and speaker identification
CAPTION IS NEAR THE PERSON SPEAKING and Speaker identified:

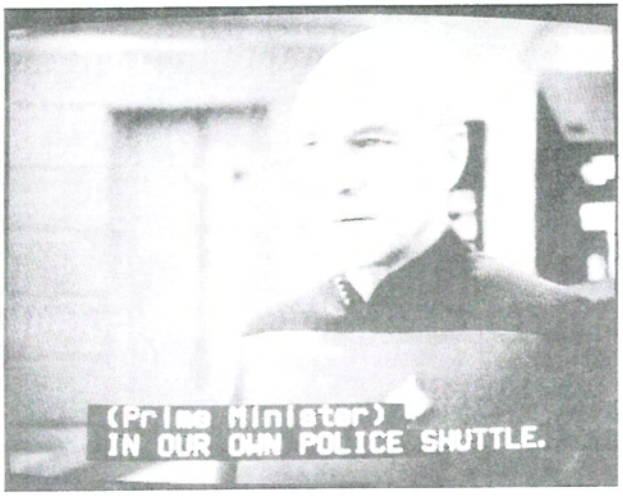
Placement and color
CAPTION IS NEAR THE PERSON SPEAKING and blue, white and yellow used

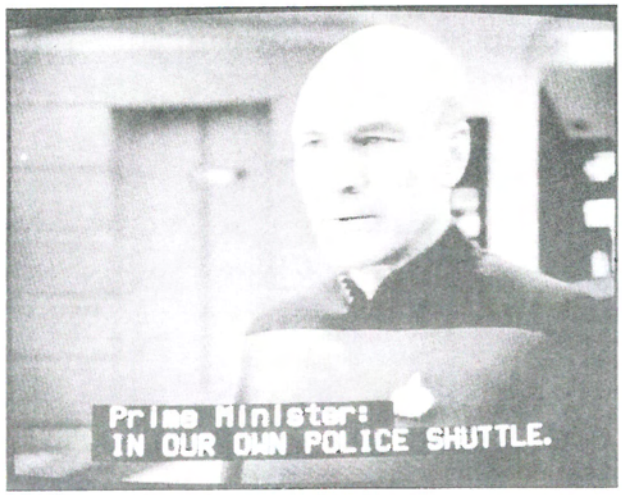
Placement only
CAPTION IS NEAR THE PERSON SPEAKING

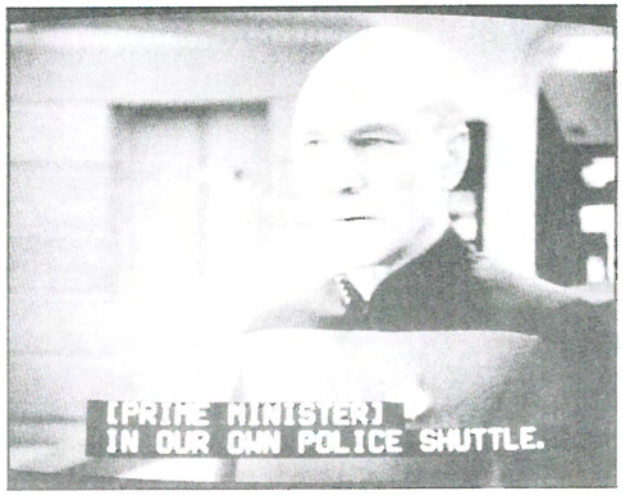
Feature	Preferred		Unacceptable	
	N	%	N	%
Placement & speaker ID	120	64	0	0
Placement & color	44	23	29	15
Placement only	25	13	15	8
Totals	189	100	44	23

SPEAKER IDENTIFICATION
MULTIPLE SPEAKERS

In this clip, there are multiple off-screen speakers.







Features used to indicate speaker:

Parenthesis
(Commander Riker)

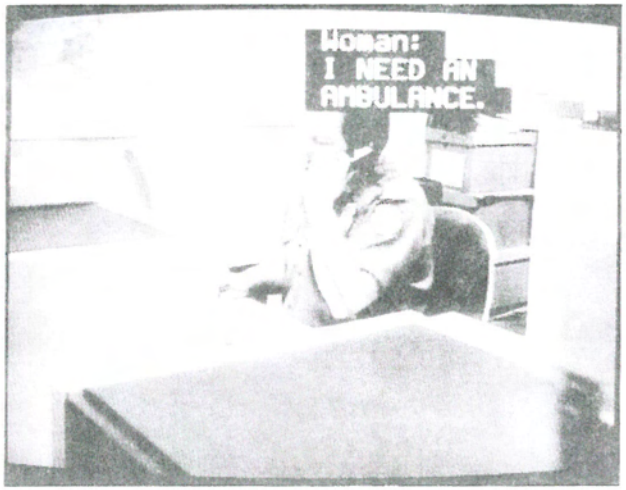
Colon
Commander Riker:

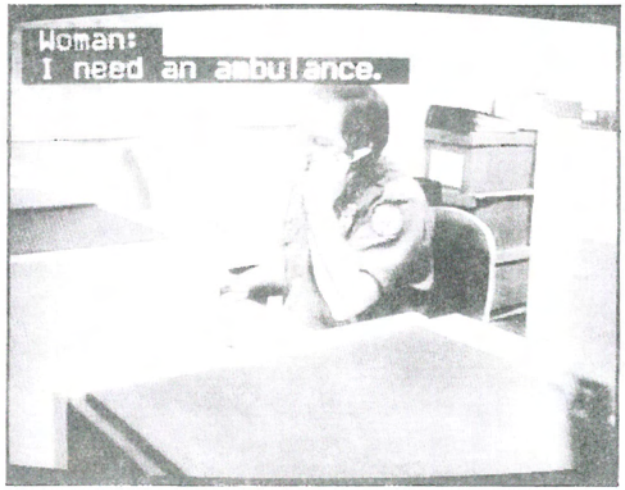
Brackets
[COMMANDER RIKER]

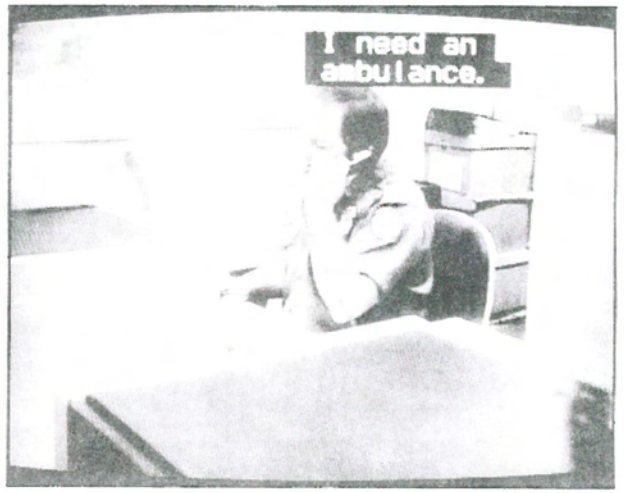
Feature	Preferred		Unacceptable	
	N	%	N	%
Parenthesis	77	41	3	2
Colon	57	30	5	3
Brackets	55	29	11	6
Totals	189	100	19	11

SPEAKER IDENTIFICATION

In this clip, a taped phone conversation is played.







Features used to indicate speaker:

Placement, speaker ID, all capital letters
Captions for operator are left; captions for woman are right
911 Operator: EMERGENCY 911 Woman: I NEED AN AMBULANCE

Placement, speaker ID, capital letters and upper/lower case letters
All captions are left
911 Operator: EMERGENCY 911
Woman: I need ambulance

Placement, capital letters and upper/lower case letters
Captions for operator are left; captions for woman are right
EMERGENCY 911I need an ambulance

Feature	Preferred		Unacceptable	
	N	%	N	%
Placement, speaker ID, all caps	78	41	7	4
Placement, speaker ID, caps & upper/lower	57	30	6	3
Placement, caps & upper/lower	54	29	9	5
Totals	189	100	22	12

Sound Effects

In a segment with several sound effects, (1) description, (2) onomatopoeia, and (3) a combination of the two were tested. In this clip, the sound effects were spaced in time so that they did not overlap. A combination of description and onomatopoeia was the preference of more consumers (56%) than was description alone (31%) or onomatopoeia alone (13%). ($\chi^2 = 53.746$, $df = 2$, $p < .001$)

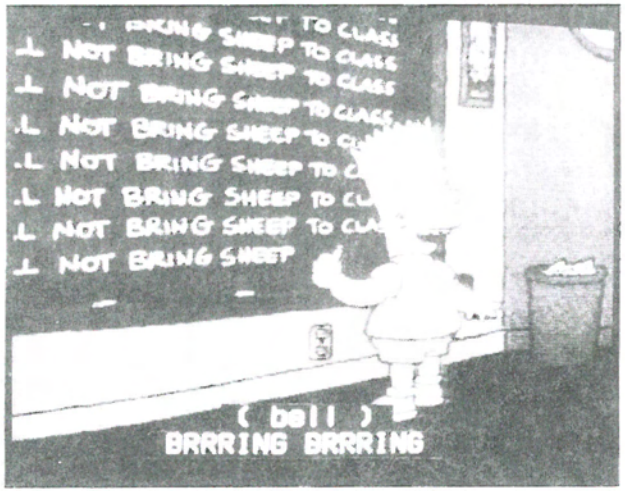
Recommendation: Where feasible, a combination of description and onomatopoeia should be used to indicate sound effects. If space or other limitations do not permit the two to be used together, descriptors should be used. Onomatopoeia should not be used alone. A descriptor is particularly important if the source of the sound effect is not obvious from the video.

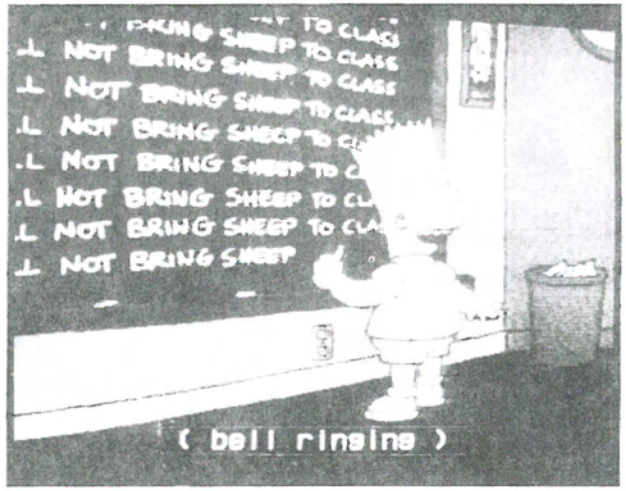
Example:

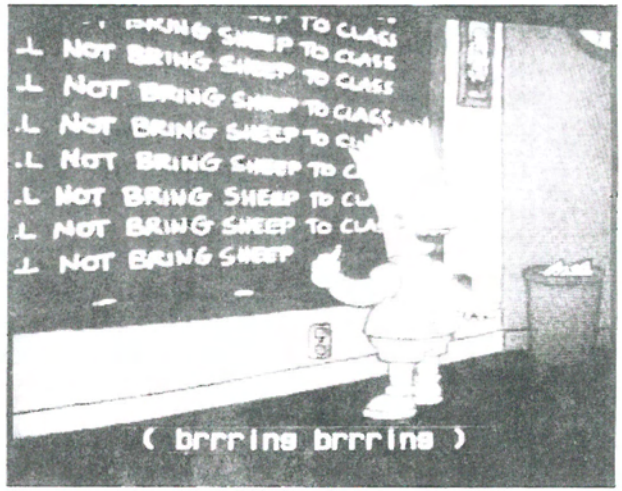
(factory whistle)
TOOOOT

SOUND EFFECTS

In this clip, there are many sound effects.







Features used to indicate sound effects:

Description & onomatopoeia
(whistle)
TOOOOOOOT

Description
(whistle blowing)

Onomatopoeia
(tooooooot)

Feature	Preferred		Unacceptable	
	N	%	N	%
Description & onomatopoeia	106	56	6	3
Description	59	31	4	2
Onomatopoeia	24	13	19	10
Totals	189	100	29	15

Music

Background music

In a sentimental scene, (1) a musical-note icon in the upper right corner (to indicate when background music was playing) was tested against (2) explicit description of the background music and against (3) no indication of background music. Description was the rather strong preference (68%) of consumers, with a small minority favoring the music icon (18%) or no indication of background music (15%). ($\chi^2 = 100.793$, $df = 2$, $p < .001$) The music icon and absence of features were also relatively often deemed unacceptable by viewers. In a scene with suspenseful music, 83% preferred description over an absence of any indication of the background music. ($\chi^2 = 80.069$, $df = 1$, $p < .001$)

Recommendation: Background music should be indicated, especially if it contributes to the plot or mood of the video. A description of the background music should be given wherever possible.

Example:

(soft, sad background music)

Singing

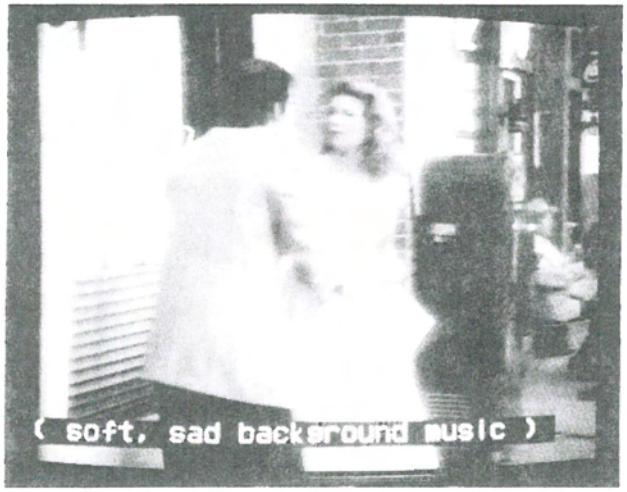
The industry convention has been to indicate singing with musical-note icons at the beginning and end of the caption. This style was tested two ways--(1) all caps versus (2) upper/lower case--and also compared with (3) a paint-on style. Some viewers commented that they did not notice the difference between all caps and upper/lower case letters. The paint-on style (a new caption feature) was rejected in this situation, and the convention of music icons combined with captions was validated. In all, 82% of respondents chose styles with the music icons. Of these, slightly more than half chose all caps, and slightly fewer than half chose upper/lower case. The differences among the three choices were significantly significant. ($\chi^2 = 21.555$, $df = 2$, $p < .001$)

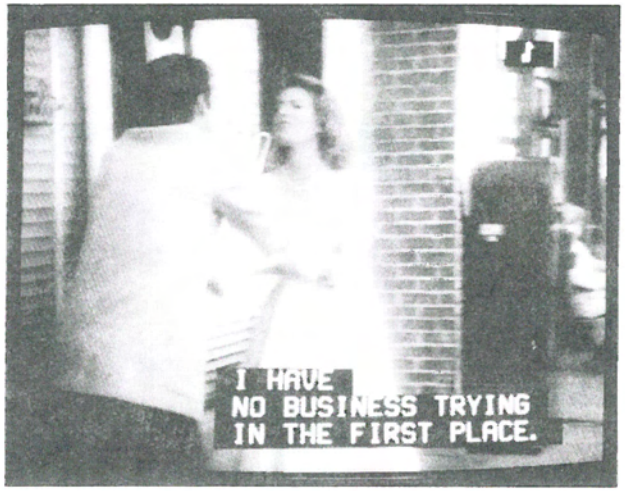
Recommendation: Continue the practice of using the musical-note icon surrounding the caption. All-caps and upper/lower-case type are equally acceptable for the caption portion.

Example: ♪ I'VE GOT A SONG TO SING ♪

MUSIC

In this clip, the background music helps to set the mood.





Features used to indicate background music:

Description
(soft, sad background music)

Music icon
music icon appears in top right corner of screen

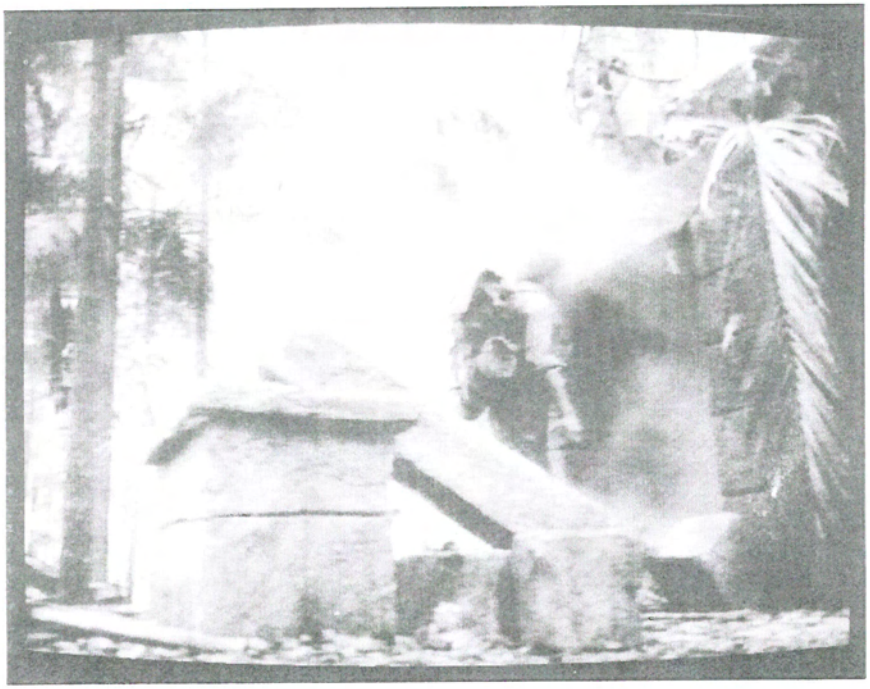
No features used
music isn't identified

Feature	Preferred		Unacceptable	
	N	%	N	%
Description	128	68	2	1
Music icon	33	18	30	16
No features used	28	15	37	20
Totals	189	101*	69	37

*Discrepancies of 1% are a result of standard rounding error.

MUSIC

In this clip, the background music alerts the viewer that something is going to happen.



Features used to indicate background music:

Description
(scary music)

No features used
Music isn't described

Feature	Preferred		Unacceptable	
	N	%	N	%
Description	157	83	2	1
No features used	32	17	27	14
Totals	189	100	29	15

MUSIC

In this clip, a woman is singing.

Features used to indicate singing:

Music icon and capital letters
♪ MUSICAL NOTE AT BEGINNING AND END OF EACH CAPTION BLOCK ♪

Music icon and upper/lower case letters
♪ Musical note at beginning and end of each caption block ♪

Paint-on lyrics
Words appear as she sings

Feature	Preferred		Unacceptable	
	N	%	N	%
Music icon & caps	80	42	2	1
Music icon & upper/lower	76	40	3	2
Paint-on lyrics	33	18	46	24
Totals	189	100	51	27

Audience Reaction

Reaction other than laughter

In a comedy scene, the live audience reacted to an actor's line. Their reaction was captioned in three ways: (1) with a flashing caption (a new feature), (2) with a standard caption, and (3) with a caption combined with identification of the source. Consumers tended strongly (77%) to prefer the descriptive version, where the audience was explicitly identified, compared to flashing captions (13%) or absence of identification of the source of the caption (11%) . ($\chi^2 = 90.063$, $df = 2$, $p < .001$)

Laughter

In a 54-second sitcom clip, the repeated laughter that characterizes sitcoms was treated with (1) a conventional descriptive caption [i.e., (audience laughing)] versus (2) a flashing caption of the words "ha ha ha," versus (3) the absence of any indication of audience laughter. Most consumers (58%) chose description as the best of these three alternatives, and few (5%) found this unacceptable. In the preferred version, the descriptor appeared seven times in 54 seconds. Only 7% preferred flashing captions, and 35% preferred no indication. ($\chi^2 = 71.841$, $df = 2$, $p < .001$)

Recommendation: Audience reaction should be captioned. This is particularly important where the reaction itself becomes part of the plot or comedy. Audience laughter should also be described. (It is of course possible that repeating the descriptor every time the audience laughs, over the length of an entire sitcom episode, would become annoying. This length of exposure was not tested. Therefore, discretion is advised; but audience laughter should be indicated much more often than is now the industry's practice.)

Examples:

(Audience)
BOO

(audience laughing)

In this clip, the audience (off-screen) is reacting to what the actor has said.

Features used to indicate audience reaction:

Identification
Audience:
WOOH

Flashing
WOOH flashes

No features used
WOOH

Feature	Preferred		Unacceptable	
	N	%	N	%
Identification	145	77	2	1
Flashing	24	13	32	17
No features used	20	11	8	4
Totals	189	101*	42	22

*Discrepancies of 1% are a result of standard rounding error.

AUDIENCE REACTION

In this clip, the studio audience frequently laughs.

Features used to indicate audience laughing:

Description
(audience laughing)

No features used
No mention of audience laughing

Flashing
(ha ha ha) flashes

Feature	Preferred		Unacceptable	
	N	%	N	%
Description	109	58	9	5
No features used	66	35	25	13
Flashing	14	7	53	28
Totals	189	100	87	46

Manner of Speaking

Emotion

In one comedy clip, an angry exchange was captioned with (1) description of the emotion, (2) exaggerated punctuation, and (3) absence of any indication of emotion in the speakers. A plurality of consumers (46%) preferred description of emotion over absence of features (preferred by 28%) and over exaggerated punctuation (preferred by 27%). ($\chi^2 = 13.746$, $df = 2$, $p < .01$) Few (5%) consumers found description unacceptable. These results were statistically significant, but were more mixed than some of the other preference data.

Recommendation: Where strong emotion is being conveyed, the emotion should be described with the caption. This feature should be used especially where the strong emotion is not entirely obvious in the facial expression and actions of the speaker. Caption writers may be concerned that this feature could be over-used. However, based on consumers' reaction, caption writers should use this feature more than is current practice.

Example:

(angrily) DAN? DAN, ARE YOU EVEN LISTENING?

Whispering

Tested were: (1) upper/lower case letters for the whispered lines, but no description; (2) description combined with all capitals; and (3) description combined with upper/lower case captions. (The description in this case was the word "whispering.") Description combined with upper/lower case was the rather strong (70%) preference of consumers, whereas upper/lower case alone was preferred by only 9% and description plus all capitals was preferred by 22%. ($\chi^2 = 115.460$, $df = 2$, $p < .001$)

Recommendation: Whispered lines should be identified as such, as follows, and combined with upper/lower case captions:

Example:

(whispering)
It's dark in here. I can't see you.

Emphasis

When a word or words are emphasized in a line, description is not possible. Three features were tested in a scene in which several significant words were emphasized. The words were indicated with (1) italic captions (2) color (a new feature), and (3) underlining. The reactions to these choices were more mixed than in many of the other items tested. The highest percentage of respondents chose italics (43%), followed rather closely by color (37%). However, color was unacceptable to a larger minority (13%) than was italics (3%). Differences in preference were statistically significant. ($\chi^2 = 17.23$, $df = 2$, $p < .001$)

Recommendation: Indicate emphasized word(s) within a caption with italics.

Example:

DON'T *EVER, EVER* SAY THAT AGAIN.

Accents

A character's southern accent was indicated (1) through description, (2) through exaggerated spelling to indicate the accent, and (3) without indication of accent. The rather strong preference (68%) of consumers was the description without an attempt to otherwise portray the accent. This method was also not controversial, with only one person finding it unacceptable. ($\chi^2 = 103.841$, $df = 2$, $p < .001$)

Recommendation: Indicate foreign or regional accent with a description at the beginning of the character's lines. (Note: This issue was tested only with a fictional character, and probably should not be generalized to other situations.)

Example:

(southern accent)
I'M SURE GLAD I DIDN'T HAVE THAT CASE.

MANNER OF SPEAKING
EMOTION

In this clip, the speakers are angry.

Features used to indicate emotion:

Description
(angrily) DAN? SAY, DAN, ARE YOU EVEN LISTENING TO ME?

No features used
DAN? SAY, DAN, ARE YOU EVEN LISTENING TO ME?

Additional punctuation
DAN?!!!! SAY, DAN, ARE YOU EVEN LISTENING TO ME!!!

Feature	Preferred		Unacceptable	
	N	%	N	%
Description	87	46	10	5
No features used	52	28	5	3
Additional punctuation	50	27	26	14
Totals	189	101*	41	22

*Discrepancies of 1% are a result of standard rounding error.

MANNER OF SPEAKING
WHISPERING

In this clip, one of the actors is whispering.

Features used for indicating whispering:

Description and caption in upper/lower case letters
(whispering)
Loren, Loren

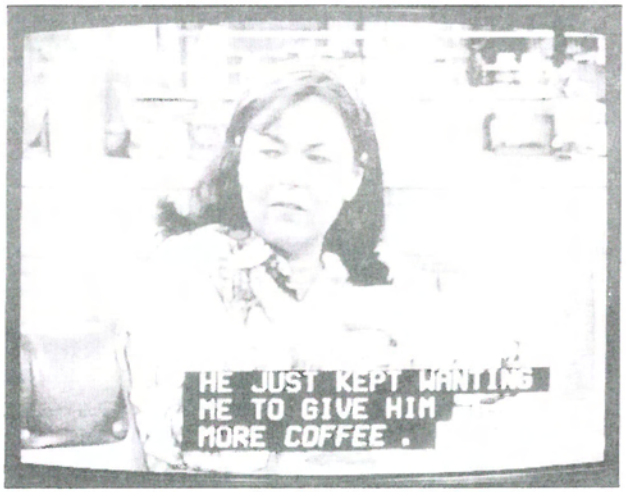
Description and caption in all capital letters
(WHISPERING)
LOREN, LOREN

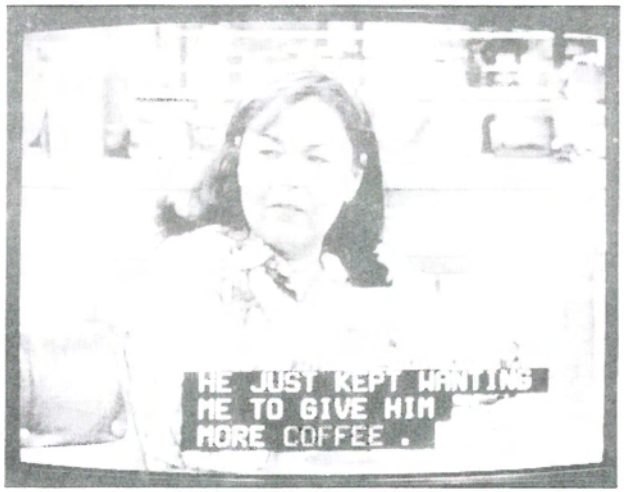
No description and caption in upper/lower case letters
Loren, Loren

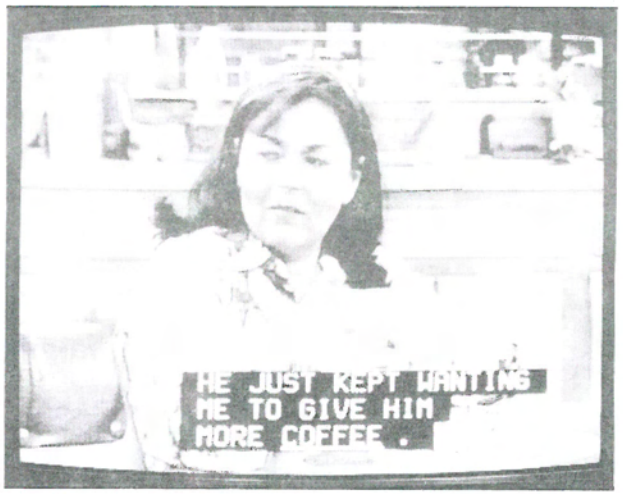
Feature	**Preferred**		**Unacceptable**	
	N	%	N	%
Description & upper/lower	131	70	1	1
Description & caps	42	22	15	8
No description & upper/lower	16	9	16	9
Totals	189	101*	32	18

MANNER OF SPEAKING
EMPHASIS

In this clip, two captions contain words emphasized by the speaker.







Features used to indicate emphasis:

Italics
COFFEE

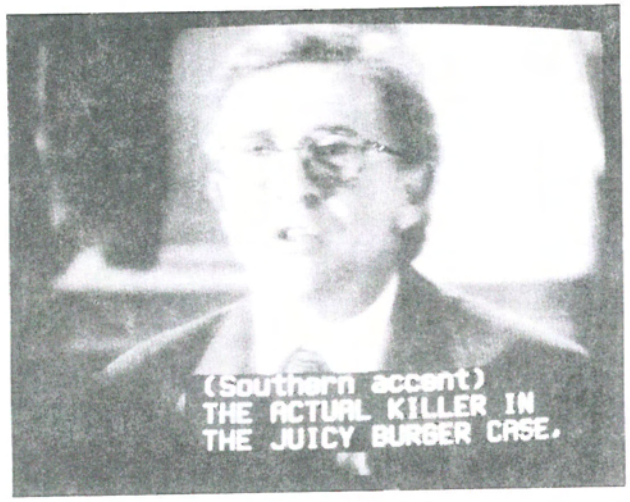
Color
COFFEE captioned in green

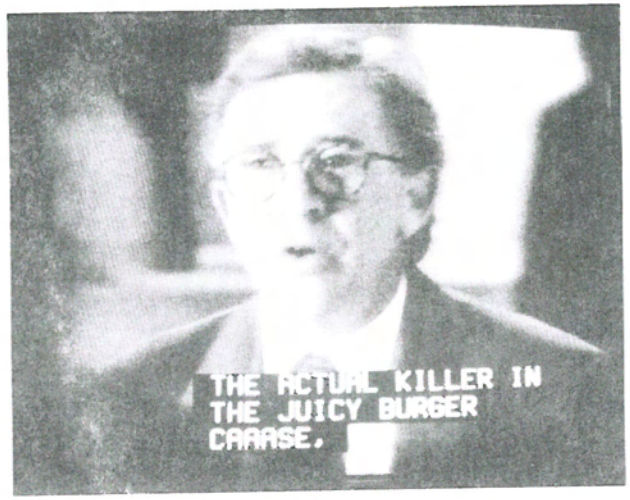
Underline
COFFEE

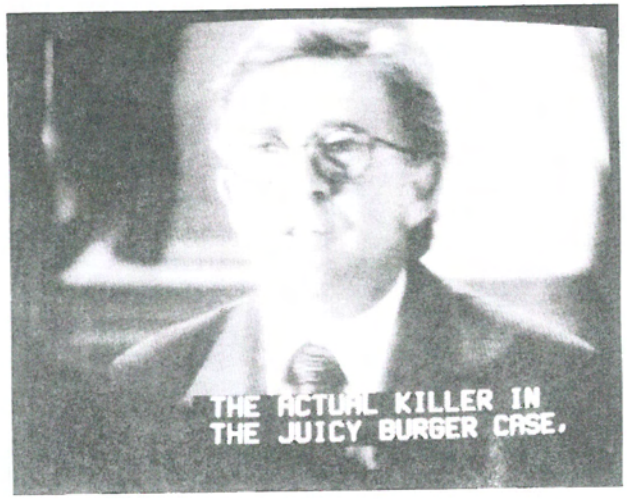
Feature	**Preferred**		**Unacceptable**	
	N	%	N	%
Italics	82	43	6	3
Color	70	37	25	13
Underline	37	20	21	11
Totals	189	100	52	27

MANNER OF SPEAKING
ACCENT

In this clip, the person speaking has a Southern accent.







Features used for indicating accent:

Description
(Southern accent)

Phonetic transcription
A RACCOOON WITH A FIIISH

No features used
No mention of accent

Feature	Preferred		Unacceptable	
	N	%	N	%
Description	129	68	1	1
Phonetic transcription	32	17	37	20
No features used	28	15	19	10
Totals	189	100	57	31

Explanation of a Pun

Puns can be difficult for deaf and hard of hearing viewers, since enjoyment of the pun tends to depend heavily on hearing. A caption (1) carrying an explanation of the pun below it in upper/lower case was tested versus (2) an absence of such explanation. The explained version was the rather strong preference, with 71% of consumers preferring it and only 1% objecting; 29% preferred a lack of explanation. ($\chi^2 = 32.196$, $df = 1$, $p < .001$)

Recommendation: Puns should be described when feasible.

Example:

I'M LOOKING FOR AMANDA HUGGENKISS
("a man to hug and kiss")

Indication of a Title

Titles have been indicated in various ways by the industry. Quotation marks, italics, and underlining were compared. Quotation marks were favored by 44%, and found unacceptable by only 1%. Italics came in second, with 29%, and underlining third, with 27%. ($\chi^2 = 10.698$, $df = 2$, $p < .01$) Underlining was unacceptable to 14%.

Recommendation: Use quotation marks when indicating the title of a book, movie, etc.

Example:

I GOT THIS CERTAIN URGE TO RENT
"HONEY, I SHRUNK THE KIDS."

EXPLANATION OF A PUN

In this clip, the conversation includes a pun.

Features used to indicate the pun:

Explanation of pun
("a man to hug and kiss")

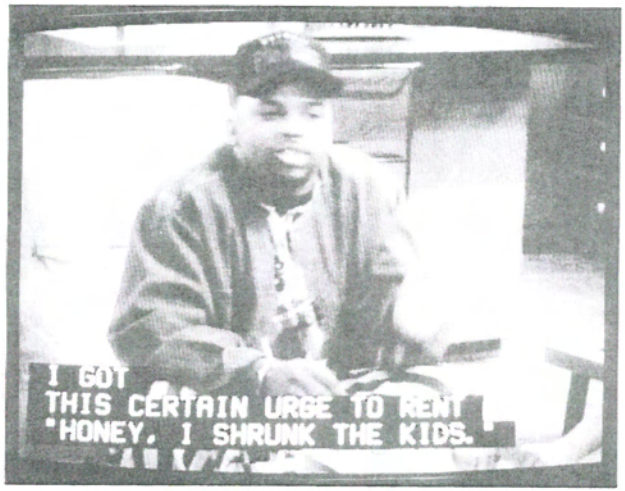
No features used
no explanation

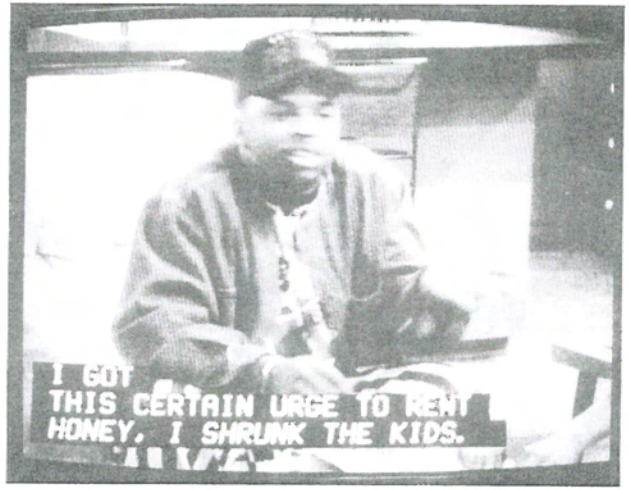
Feature	Preferred		Unacceptable	
	N	%	N	%
Explanation	133	71	2	1
No features used	55	29	5	3
Totals	188*	100	7	4

*One person did not respond to this item.

PART OF LANGUAGE
TITLE

In this clip, the person speaking says the title of a movie.





Features used to indicate movie title:

Quotation marks
"HONEY, I SHRUNK THE KIDS"

Italics
HONEY, I SHRUNK THE KIDS

Underline
HONEY, I SHRUNK THE KIDS

Feature	**Preferred**		**Unacceptable**	
	N	%	N	%
Quotation marks	84	44	2	1
Italics	55	29	3	2
Underline	50	27	26	14
Totals	189	100	31	17

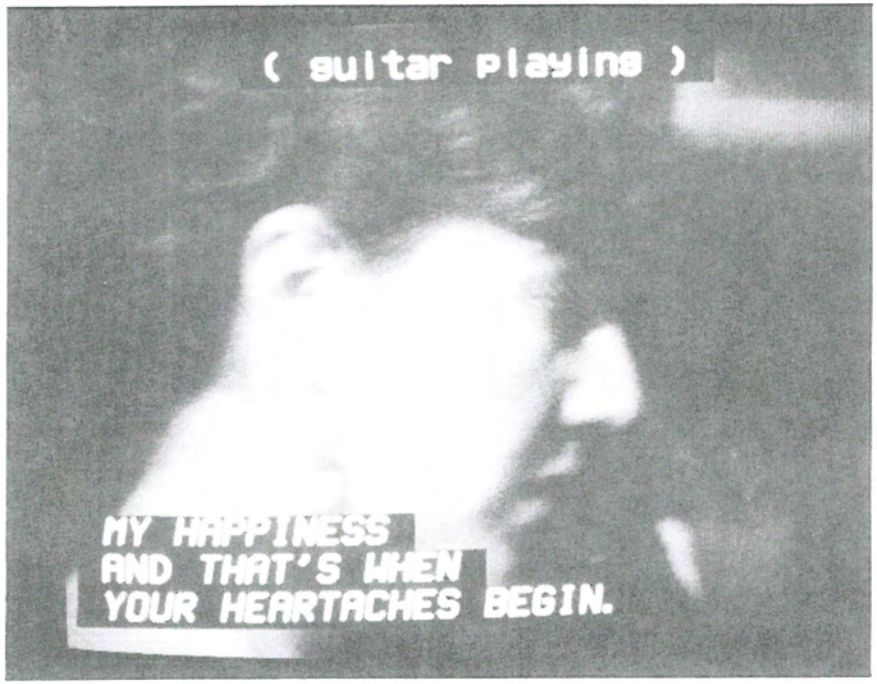
Quantity of Indicators

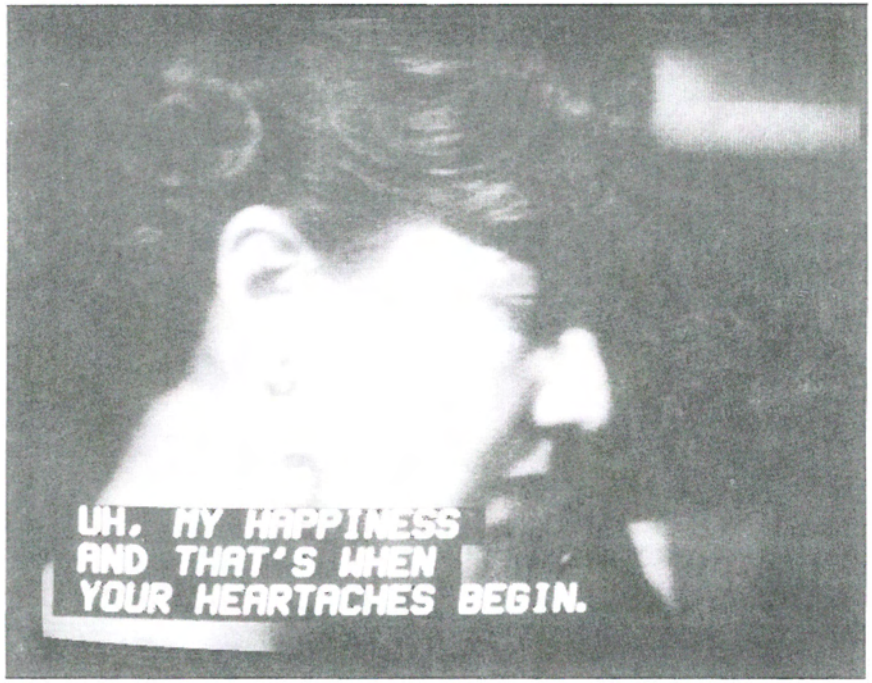
Two possible constraints to indicating non-speech information are reading rate and complexity of the visual environment. When is there too much information? One set of clips was used as a beginning test of where "too much" may be a factor in consumer preference. A clip with a busy audio background (with sounds carrying information) was captioned at three reading rates: approximately 149 words per minute, 167 words per minute, and 192 words per minute. The difference in reading rate was a factor of the amount of information about background sounds. The fastest, most complex version was the most preferred: 51% preferred this version, 35% preferred the middle version, and only 13% preferred the version with no indications of non-speech information (and a roughly equal number of people found this version unacceptable). ($\chi^2 = 41.523$, $df = 2$, $p < .001$)

Recommendation: If a descriptive caption or feature would in any way clarify or enhance the viewer's experience of being connected with the audio, it should be indicated. Consumers prefer that more of such information be included than is now the case.

AMOUNT OF NON-SPEECH INFORMATION TO INCLUDE

In this clip, different amounts of non-speech information are captioned, in addition to the dialogue.





Features used to indicate amount of non-speech information:

Dialogue and all sounds captioned

Dialogue and some sounds captioned

Dialogue and no sounds captioned

Feature	Preferred		Unacceptable	
	N	%	N	%
All sounds	97	51	8	4
Some sounds	67	35	1	1
No sounds	25	13	24	13
Totals	189	99*	33	18

*Discrepancies of 1% are a result of standard rounding error.

References

King, C. M., LaSasso, C. J. & Short, D.D. (1994). Digital captioning: Effects of color-coding and placement in synchronized text-audio presentations. In J. Willis, B. Robins, & D.A. Willis (eds.), *Educational Multimedia and Hypermedia, 1994* (pp. 329-334). Charlottesville, VA: Association for the Advancement of Computing in Education.

King, C. M. and LaSasso, C. J. (1993). Caption format research project. Presentation for Electronic Industries Association Television Display Standards Section. Washington, D.C., May.

Appendix A

Questionnaire on Awareness of Features

Code Number: N=189

LOOK AT THE CAPTION.

Correct answer - 18%
Partially correct answer - 13%
Incorrect answer - 16%

In this caption, what does >> mean?

Please check ALL POSSIBLE answers:

	28%	The speaker began a new topic
*	31%	A different speaker began talking
	1%	The speaker began to talk louder
	2%	It is a mistake in the captioning
	53%	Don't know/not sure

* correct answer

LOOK AT THE CAPTION.

Correct answer - 4%
Partially correct answer - 27%
Incorrect answer - 22%

In this caption, why are the letters *SLANTED*?

Please check ALL POSSIBLE answers:

	18%	The person is shouting or yelling
	3%	The person has an accent
*	14%	The person is thinking those words, not speaking those words
*	21%	The person speaking is not on the screen
	9%	It is a mistake in the captioning
	47%	Don't know/not sure

* correct answers

LOOK AT THE CAPTION.

Correct answer - 25%
Partially correct answer - 3%
Incorrect answer - 20%

Upper case = LETTERS ARE IN CAPITALS (like most captions)
Lower case = letters are small, like this
Upper/Lower case = Some are capitals, some are small, as in a book or magazine

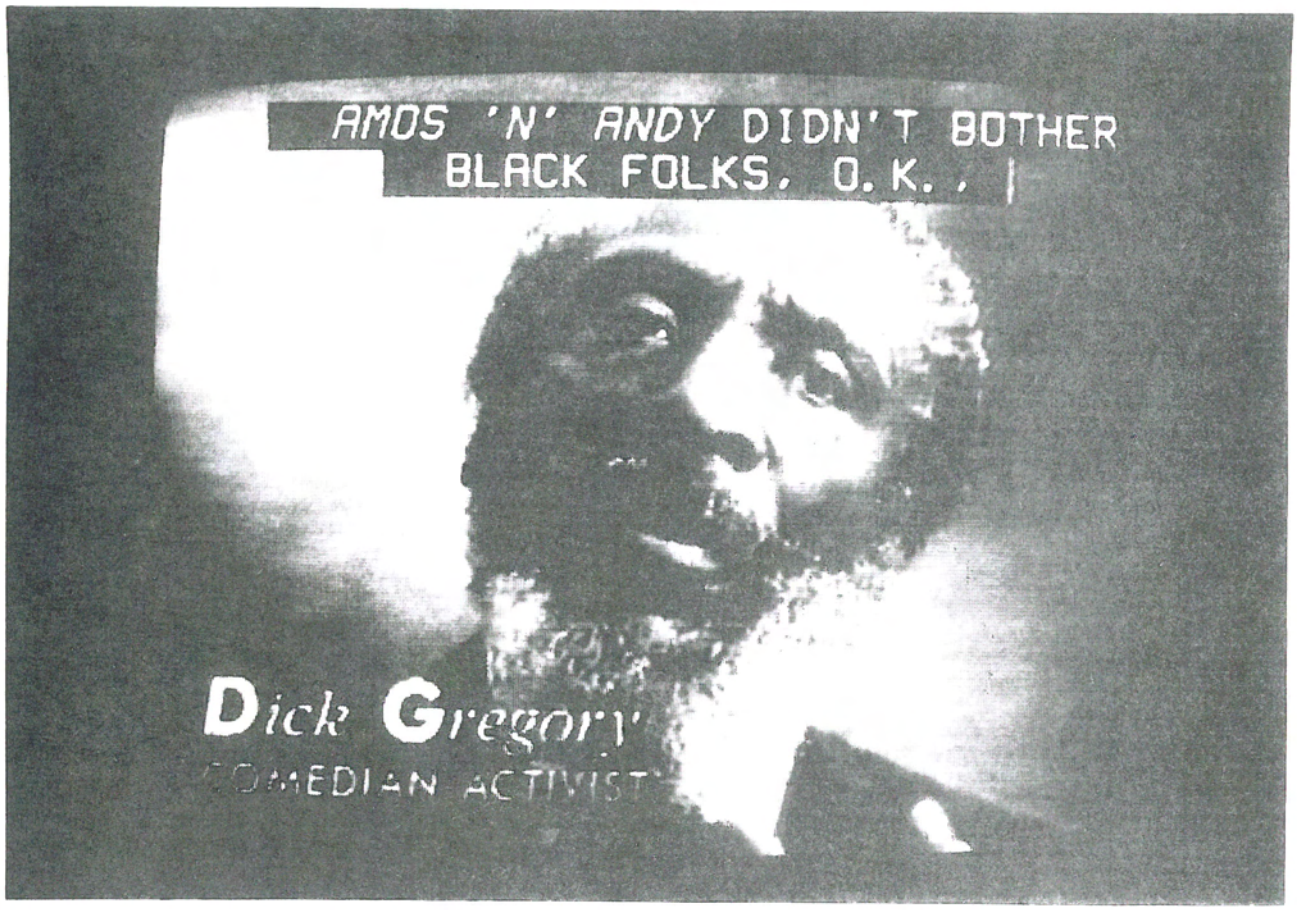
In this caption, why are the letters upper/lower case?

Please check ALL POSSIBLE answers:

	12%	The person speaking is not on the screen
*	29%	The person is whispering
	3%	The person has an accent
	4%	The person is shouting or yelling
	4%	It is a mistake in the captioning
	52%	Don't know/not sure

* correct answer

LOOK AT THE CAPTION.



Correct answer - 65%
Partially correct answer - 0%
Incorrect answer - 7%

In this caption, why are a few words *SLANTED*?

Please check ALL POSSIBLE answers:

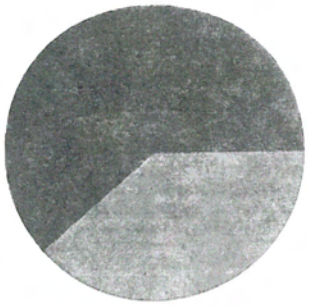
	4%	The person is shouting or yelling those words
	2%	The person is whispering those words
*	65%	Those words are a title (like a book, movie, TV show, etc.)
	2%	It is a mistake in the captioning
	28%	Don't know/not sure

* correct answer

Appendix B

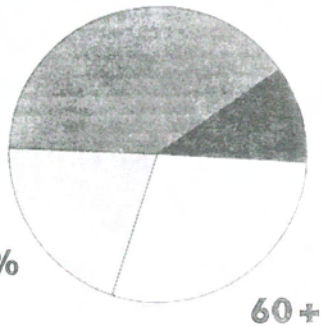
Demographic Characteristics of the Sample

Gender



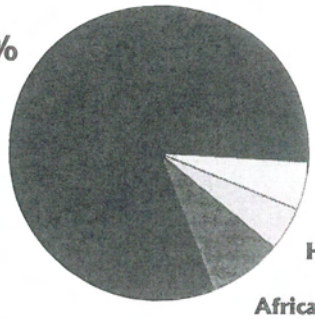
N=189

Age



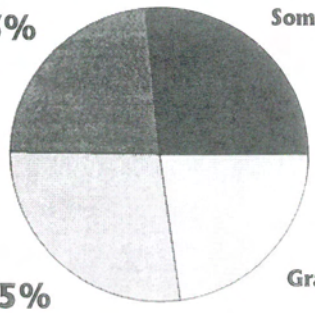
N=189

Ethnicity



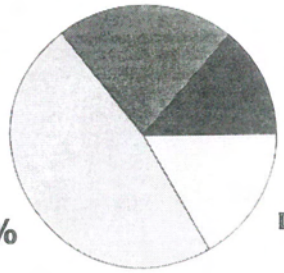
N=189

Highest Level of Education



N=189

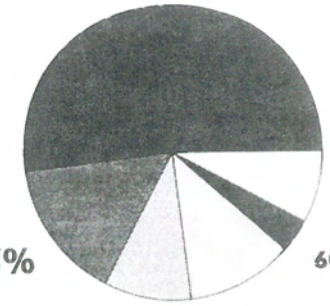
Level of Hearing Loss in Better Ear



N = 189

Self-reported estimates of hearing loss

Age at First Occurrence of Hearing Loss



N = 189

Appendix C

Companies and Organizations that Responded to the Draft Guidelines

Companies and Organizations that Responded to the Draft Guidelines

John P. Armour
Captioner and Media Producer
920 Stryker Avenue
West St. Paul, MN 55118

The Caption Center
125 Western Avenue
Boston, MA 02134

The Caption Company
1700 Water Place
Suite 300
Atlanta, GA 30339

The Caption Connection
1320 Edgewater NW
Suite B10
Salem, OR 97304

CaptionMax
708 North First Street
Suite 131
Minneapolis, MN 55401

CBS Inc.
51 West 52 Street
New York, NY 10019

Closed Captioning Services
2215 Oak Industrial Dr., N.E.
Suite 18
Grand Rapids, MI 49505

Maryland Public Television
11767 Owings Mills Boulevard
Owings Mills, MD 21117-1499

National Association of the Deaf
Captioned Films/Videos Program
Woodland Executive Center, Suite I
1218 Reidville Road
Spartanburg, SC 29301

Self-Help for Hard of Hearing People
7910 Woodmont Avenue
Suite 1200
Bethesda, MD 20814

Appendix D

Advisors, Consultant, and Staff

Advisory Committee

Rosemary Bennett
Gallaudet University

Marjorie Boone
Self Help for Hard of Hearing People

Gertrude Galloway
Marie Katzenbach School for the Deaf

Michael Grossman
WGBH/The Captioning Center

Carl Jensema
Institute for Disabilities, Research & Training, Inc.

Jeff Hutchins
VITAC, Inc.

Sue Mather
Gallaudet University

JoAnn McCann
Department of Education

Linda Randall/Len Helmniak
National Captioning Institute

Ramon Rodriquez
U.S. Department of Education

Toby Silver
Television for All Coalition

Patti Singleton
Gallaudet University

William Stark
National Association of the Deaf

Consultant: Ves Bennett
The Ganymede Group, Inc.

Technology Assessment Program Staff on this Project

Judith E. Harkins, Director
Ellie Korres, Research Associate
Beth R. Singer, Research Associate
Barbara M. Virvan, Research Associate
Yoon K. Lee, Video Production Specialist

Gallaudet University, in Washington, D.C., is the world's only liberal arts university for deaf students. In addition to offering on-campus educational programs from the preschool to doctoral level, Gallaudet is an internationally recognized center for research, program development, and consultation related to deafness and hearing loss. Gallaudet University is an equal opportunity employer/educational institution and does not discriminate on the basis of race, color, sex, national origin, religion, age, hearing status, disability, covered veteran status, marital status, personal appearance, sexual orientation, family responsibilities, matriculation, political affiliation, source of income, place of business or residence, pregnancy, childbirth, or any other unlawful basis.