McMASTER
UNIVERSITY

MICHAEL G. DeGROOTE
SCHOOL OF BUSINESS

RESEARCH AND
WORKING PAPER
SERIES

INTERFACE SUPPORT FOR INFORMATION SEARCH IN DECISION MAKING: EFFECTS OF VOICE AND TEXT OUTPUT MODES WITH INFORMATION ABSTRACTION

By

N.P. Archer, M.M. Sigmund, J.P. Wollersheim, and Y. Yuan
Michael G. DeGroote School of Business
McMaster University,
Hamilton, Ontario, Canada L8S 4M4

Working Paper # 390

January, 1994

TER UNIVERSITY
n Street West
., Ontario, Canada L8S 4M4
e (416) 525-9140

INTERFACE SUPPORT FOR INFORMATION SEARCH IN DECISION MAKING: EFFECTS OF VOICE AND TEXT OUTPUT MODES WITH INFORMATION ABSTRACTION

By

N.P. Archer, M.M. Sigmund, J.P. Wollersheim, and Y. Yuan
Michael G. DeGroote School of Business
McMaster University,
Hamilton, Ontario, Canada L8S 4M4

Working Paper # 390

January, 1994

Interface Support For Information Search In Decision Making:
Effects Of Voice And Text Output Modes With Information Abstraction

N.P. Archer, M.M. Sigmund, J.P. Wollersheim, and Y. Yuan

School of Business
McMaster University
Hamilton, Ontario L8S 4M4
Canada

Interface Support For Information Search In Decision Making: Effects Of Voice And Text Output Modes With Information Abstraction

Abstract

This paper describes a human-computer interface which was designed to study the effectiveness of voice and text output modes, along with information abstraction, in decision making. The interface was tested in an exploratory study by population samples of both domain experts and non-experts in an apartment choice problem. We observed that there was no difference in user preferences for both user communities between text and voice or between text and text plus voice. However, text plus voice was preferred over voice alone. The speed of interaction with a pure voice interface was slower than either of the others. The choice of output mode, although it affected time efficiency, had no effect on the decisions made. We also observed that domain experts used higher level information abstraction more than detailed information, in contrast with non-experts, and heuristic experts were faster than analytic experts in completing decision tasks. An entropy calculation showed that domain experts were also significantly less random in their choice of information attributes. Our findings indicate that text or a combination of voice plus text can be effective output modes, and that information abstracted at several levels can be useful for certain types of users.

1. Introduction

Recent advances in multimedia computer equipment and software (Bly et al 1993; Hodges & Sasnet 1993) have created many opportunities for developing more usable systems to support the decision-making process, as these systems are rich in expressive and interactive ways in which information can be presented and used by decision makers. However, there have been relatively few studies into how such systems can most effectively support the decision process or how they may impact decision outcomes. Multi-media computing can add many dimensions to the characterization of information, with the potential addition of voice, video, animation, and images to the usual text and graphics output modes, and voice and gesture to other more frequently used input modes (keyboard, mouse, touchscreen, etc.). Since these dimensions create many different ways with which decision makers can be presented with information, there is currently much uncertainty in how this should be done. There are a number of different objectives which computer interface designers need to consider when choosing the presentation mode. These may include minimizing effort, maximizing efficiency, matching user preference for presentation mode, and providing information in such a manner that it does not result in a mismatch with the user's individual characteristics, such as experience and cognitive style. As one would expect, it is not possible to meet all of these objectives simultaneously, and research is needed to determine which alternatives are suitable in particular decision-making situations.

Our study focuses on a comparison of user performance and preferences with speech and text output modes, combined with an examination of the impact of information abstraction support at different levels in these modes. We also measure the impact of individual differences on performance and preferences, but the objective of our research is not to find particular interface characteristics which match well with individuals who have certain individual characteristics, but to determine how interfaces can be designed to assist users in acquiring decision information, across a range of user characteristics and preferences.

In the following, we first review previous work on text and speech output, and then discuss abstraction and domain experience, and how certain user characteristics can be measured and used in attempting to explain user performance. Then the main experiment is described, where an exploratory study analyzes the performance and preferences of 48 task domain experts and 22 domain non-experts in terms of user cognitive style, how information abstractions are used, and the type of information output mode used. The computer interface used in the experiment is described, and then the data analysis is discussed. Finally, our conclusions and findings are detailed.

2. Dimensions Of The Study

2.1 Voice And Text Output Modes

There has been some research into the effects of using either voice or text or combined voice and text. Streeter (1990) outlined the advantages and disadvantages of speech in comparison with text. Nugent (1982) and Baggett and Ehrenfeucht (1983) found that a dual modality output presentation tended to give subjects better comprehension and retention than single modality outputs. DeHaemer and Wallace (1992) suggest that, based on existing research results, the visual and aural modes of receiving information appear to be non-interfering and may enhance performance for certain tasks. They observed the effect of voice output on computer-supported decision making, where voice instructions were used to solve a visual decision problem, and found an interactive effect between user decision style and the use of computer synthetic voice. Chalfonte et al (1991) compared voice and text annotation in co-authored documents in terms of interactivity and expressiveness, and found that voice was preferred for addressing higher level issues in suggesting document modifications, but text was preferred for more detailed and lower level comments. We believe that more study is needed to determine the usefulness of voice as an information medium in its own right, as well as a supporting or complementary role, all of which are possible now with multimedia systems.

2.2 User Experience

Nielsen (1993) indicates that one user characteristic, experience, has three main dimensions: experience with the system, with computers in general, and with the task domain. All of these types of experience may have an impact on how users utilize data presented through a computer interface. In our experiments we held the first two of these factors constant by selecting subjects who were relatively familiar with computers, and all were equally unfamiliar with our relatively simple interface because none had used it before. However, we were interested in task domain experience, since this is known to have effects of particular interest to us. For example, Archer and Kao (1993) found that users with domain experience were much more likely to make use of high level abstractions in problem solving. Batra and Davis (1989) also found that domain experts focused on generating an holistic understanding of a problem before solving the problem, but novices tended to have an inability to map parts of the problem description into appropriate knowledge structures. The ability to reason about a problem depends upon previous experience in that domain, since this provides a framework or schema with which to structure known information (Staggers and Norcio 1993).

2.3 Abstraction Levels

It is useful in reasoning about a problem, if related information is presented to a user in the form of a model which is an abstraction of reality (e.g. a database record is an abstraction of the entity which it represents). But users have limited short term memory, which can handle only about seven "chunks" of information (Miller 1956). This makes it important for decision makers to have a choice of model views at different levels of abstraction, with the higher levels containing fewer and more generalized information chunks, to make more general considerations easier to assimilate. Information abstraction is widely used in many forms to reduce complexity in information acquisition and problem solving (Ossher 1987). It can help users to focus on certain facets of the problem, to deal with the problem at a desired level of complexity, and to think about the problem rather than being occupied with the details. Information abstraction finds application in diverse areas such as problem solving (Anderson 1985), formulating strategic problems (Ramaprasad and Mitroff 1984), systems design (Guindon et al 1987), and simulation modeling (Bond and Soetarman 1988).

2.4 Cognitive Style

Another frequently measured user characteristic is cognitive style. Benbasat and Taylor (1978) reviewed the impact of cognitive style on management information systems design, although the impact of cognitive style on user performance seems to be considerably less than the impact of task type and decision situation (Huber 1983). One category of cognitive style is "thinking mode" (Bariff and Lusk 1977). Taggart and Robey (1981) suggested that the T/F and S/N scales from the Myers-Briggs Type Indicator (MBTI) (Myers and McCaulley 1985) instrument can be used together to provide four classifications of thinking mode: ST, NT, SF, and NF, when the indicated characteristics are dominant. They classified decision makers with ST scores as having an analytic decision-making style, while those with high NF scores have an heuristic decision-making style. The NT and SF types are an intermediate or neutral classification. De Haemer and Wallace (1992) found that these classifications were significant in explaining the results of a decision task experiment involving computer voice. O'Keefe and Pitt (1991) found weak evidence that preference for display type can be partially explained by cognitive style. Davis, Barnes, and Jackson (1992) suggested that the method of communicating information to users will be more effective if it is matched to their cognitive style. Blaylock and Rees (1984) concluded that cognitive style influences a decision maker's evaluation of an unstructured, strategic planning problem.

3. The Experimental Interface

During this study, we did not attempt to measure the quality of decisions made by users, but concentrated on the information acquisition process. To evaluate this process and user interface preferences, a number of subjects were asked to use a computer interface to assist them in performing an apartment selection task similar to that introduced by Payne (1976) in the form of information attribute values printed on cards on an information board, and in a modified computerized form by Todd and Benbasat (1993). This task type is suitable because it is a problem which is well-known to almost all university students, the task domain expert population in our case. This task was set up so that each alternative selection had the same attributes in which subjects might or might not be interested, when making a decision about choosing such an apartment. Information attributes were both qualitative and quantitive, of the type normally used in making apartment choices. The displays included fourteen attributes of common interest at the Specific (most detailed) level of abstraction, four attributes at the General (intermediate) level of abstraction, and one attribute "Overall" at the highest abstraction level. The Appendix shows the attribute values for one of the apartments used in the study. The interface displayed to users is shown in Figure 1. Since this display gives equal access to all information attributes in a particular apartment, it avoids hampering access to information with the resulting bias towards selecting attributes at particular abstraction levels which may occur in an hierarchical interface such as a menu system. Information was not displayed unless requested by the subject, and then only for a limited time. This information retrieval process was recorded by the system for later analysis. Since apartment selection is a matter of individual preference, there were no right or wrong answers in the exercise. In our experiments, we used a constant number of four alternative apartments in each decision situation.

*** Insert Figure 1 about here ***

The interface was designed with the aid of the Asymetrix Toolbook© software construction environment, running under Microsoft Windows©3.1 on a 486 PC. Recorded female voice output was supported by a Creative Labs Soundblaster© unit. Subjects used a mouse to select screen buttons which would give information on particular attributes of the apartment in question, basically simulating a database browsing interface. The system recorded and time stamped on a text file each button chosen by each user. These files and computerized questionnaire responses were used later for statistical analysis.

Subjects were initially given a short animated demonstration (using both Text and Voice) on how to use the system to search for information and to adjust scoring preferences for the apartments. Then they were given a simplified problem with two apartments where they could learn directly how to use the interface, before moving on to the first of two "Scenes" or apartment selection tasks. A "Scene" for this experiment is defined as a combination of four apartments (see Figure 1), among which the subject was asked to make a choice and to rank the apartments according to his or her own personal preference. For each Scene, subjects used one of the output modes (Voice, Text or Both voice and text), which were controlled by the experimenter. The total time required by a subject to complete an apartment selection experiment varied from about 20 minutes to 45 minutes.

A scoring button shown beside each of the four Apartment choices in Figure 1 could be clicked at any time by the subject, to increment or decrement the score of any apartment between the values of 0 and 10, as a memory aid and as a means of "zeroing in" on a choice of apartment. The final score assigned to each apartment was used to indicate the ranking of the apartment, with a higher score indicating a more desirable apartment. An additional memory aid was the black background colour for the data attribute buttons. This background colour changed to gray when the button had been selected by the subject, serving as a reminder of which information had been selected, but did not restrict the subject from returning to that information at any future time. No other memory aids were supplied.

The time taken to play back information from any of the buttons when in Voice mode was used as a standard, so if Text or Both (Voice and Text) modes were used, exactly the same time was used to display Text information as the equivalent voice output. Text information was shown in a small pane at the upper right of the screen (see Figure 1). Either voice or text output could be cut off during the display time by clicking on any other button in the display. When the subject completed analyzing data and scoring the apartments, the Final Scoring button could be clicked to leave the apartment selection process. The subject was then given a final opportunity to adjust the relative scores before going on to the questionnaire which followed the decision making task.

4. Experiments

Since it was anticipated that task domain experts (people who had previous experience in searching for and selecting suitable living accommodation) and domain non-experts (people with no previous apartment hunting experience) would interact differently with the interface, subjects from both classifications were used in the experiments. The task domain experts were 48 MBA students; the median number of times they estimated they had conducted such a search was 6.5. Almost all said they had searched for living quarters at some time in the previous year. The task presented during the experiment was to select their best choice of apartment in the university environs from among those presented in a particular scene (four alternative apartments). 21 of these subjects were female and 27 male; each was paid $10 for taking part in the experiment. MBTI results revealed that, in this group, there were twelve Analytics, five Heuristics, and thirty-one Neutrals, according to the Taggart and Robey (1981) classification structure. All were reasonably familiar with the interactive use of computers.

There were 22 subjects who were not task domain experts. These were chosen from a class of senior high school students. None of these subjects had previously either taken part in, or been taught how to do, searches for living quarters. There were 14 females and 8 males. The subjects were each paid $5 for their time. MBTI results indicated that there were seven Analytics, six Heuristics, and nine Neutrals in this group. This group attended high school in a town near the university, so were familiar with the university environs. Subjects from this group were asked to assume that they were planning to attend the university in the coming year, and would need to find apartment accommodation at that time. All were reasonably familiar with the use of personal computers.

The main experiment concerned the 48 task domain experts. This was a repeated measures unbalanced within-subject design. Each subject compared one of the three possible pairs of output modes (Voice, Text, and both voice and text (Both)) by carrying out two apartment selection tasks using two different output modes. Each paired comparison used two different scenes of four apartments each. Although the apartments all had the same set of attributes, the attribute values used were varied among the apartments within realistic limits. Thus the overall task complexity was held relatively constant across the total of four scenes with four apartments each, among the three different output mode comparison pairs. Four subjects used each scene pairing for each combination of output modes, with two of these subjects doing the task with the output modes in reverse order to the other two subjects, so the impact of interface learning effects could be determined and balanced out.

The domain non-experts used the same interface in the same manner, but this was a subsidiary experiment and was not a complete evaluation of the three interface pairings.

5. Data Analysis

This experiment differs from previous experiments of this nature (Payne 1976; Todd & Benbasat 1993) in that, although it was simple to access any of the required information at will, the information was only displayed for a limited time after it was accessed in order to achieve a fair comparison between voice and text modes. Previous experiments have displayed information continously after it was accessed, and protocol analysis was used to determine the thought processes of the subjects. Each subject was observed throughout our experiments, and their activities were recorded by the computer which captured and time stamped each button press.

We define the ButtonTime for a particular button as the time difference between the time that button was pressed and the time the immediately following button was pressed, summed over all button presses for that particular button during a scene (four apartments). ButtonNo is defined as the total number of times a particular button was pressed (thus accessing that particular data attribute) during a scene. We also define the Master time for a particular button as the time taken to display the information when that button was pressed. For all modes (Voice, Text, and Both), each particular apartment attribute value was always displayed the same length of time, (defined as the Master time), as the amount of time for the equivalent voice message to be played back. Since it was possible for the subject to move on to another button before the end of the Master time interval, we also define a variable ButtonCutoff as the percentage of the total Master time for that button that the subject allowed before pressing the next button, for all presses of a particular type during a scene.

We define the SceneLength as the sum of all ButtonTimes over all buttons pressed during a scene analysis, excluding any scoring buttons. Similarly, the number of times each button was pressed was calculated, and the SceneTotalNo is defined as the total of all button presses during scene analysis, excluding scoring ButtonTimes. The SceneCutoff is defined as the actual display time for the non-scoring buttons used in the scene divided by the Master display times for these buttons, expressed as a percentage.

5.1 Relative Importance And Focusing On Information Attributes

Table 1 shows rankings by both experts and non-experts for the information attributes used in the experiments, both by subjective rating in questionnaires and by the actual number of references.

*** Insert Table 1 about here ***

Considering only the domain expert data, of the 19 possible attributes which could be queried for each apartment, only two (Overall and Rate) were queried on average more than once per apartment (averages greater than 4.0). The ranking of the "Overall" attribute was lower for the non-experts, although they also queried Rate most frequently of all the attributes. The Spearman correlation coefficient between rankings for the number of references between experts and non-experts revealed an r_s of 0.78 ($p<.01$) which does indicate a significant association between the rankings by experts and non-experts. From the table, the Rate attribute was clearly regarded as the most important by both groups, and such attributes as Closet, Shopping, and Parking were ranked at the low end of the scale. However, both groups perceived the Overall attribute as having considerably less importance than their retrievals of this attribute actually indicated. This may be because this attribute was used both as a source of information and as an anchoring indication by some of the users of the "overall" characteristics of the particular apartment. The attributes at the intermediate abstraction level (Environment, Rental, Location, and Interior) achieved relatively low usage rankings for expert users, although they were closer to the average rate for the non-experts (see Section 6.5). These results give a message which is critical to successful interface design - the importance attached to information by the user may not match the way information is organized into the hierarchy. There are two solutions to this problem: 1) re-organize the hierarchy, or 2) make it convenient for the user to access any information attribute easily and directly.

We would expect that the domain experts would be more selective in their choice of data, because of their experience in analyzing such situations. This would show up in a greater tendency to access certain attributes, and can be evaluated in terms of information entropy H. This is a measure of randomness in information retrieval, and can be defined as (Shannon 1948)

$$H = -\sum_{i=1}^{i=n} p_i \ln p_i$$

A probability p_i is associated in this experiment with the likelihood that a particular information attribute i will be selected by a subject. An estimate of p_i is the number of times attribute i is actually accessed during a scene, divided by the total number of attribute accesses. If the subject accesses each of the nineteen attributes strictly by chance, then the appropriate probability estimate that a particular attribute will be accessed is 1/19, giving a result of $H = 2.944$ for entirely random information accesses. This maximizes H, the randomness in information retrieval. On the other hand, if an individual always accessed only one particular attribute, then there is no randomness and $H = 0$, its minimum value. We would expect H to be smaller for experts than non-experts, because experts would be more likely to be selective in accessing certain information attributes. This turns out to be the case. The median value of H calculated for the 22 non-experts was 2.84, while for the 48 experts it was 2.68. A one-tailed Mann-Whitney test of the difference between these two groups was significant ($p=.011$).

5.2 Output Mode: Preferences And Impact On Decisions

Figure 2 lists the questions from the comparison questionnaire which the subjects completed after both interfaces had been evaluated. Subjects were requested to indicate the output mode preferred. The statistical analysis of the responses from the expert subjects is shown in Table 2. It is very clear from this analysis that there is no significant difference in preference between the Voice or Text output modes. However, in the Voice versus Both comparisons, Both is significantly preferred to Voice for all eight questions. When Text is compared to Both, the users did not display a significant preference for either, except that Text was preferred in terms of speed (Question 4). These relative preferences appears to imply that Voice serves as a supporting mode when used in conjunction with Text, so both together are readily preferred to Voice alone. In our experiments, the voice output contained exactly the same information as the text output, so the voice output duplicates the text output rather than serving as a complementary output as in some multimodal applications (for example, voiceover as used to complement graphics, video, or images in most television advertising). However, the

Text mode when used by itself appears to be on an equal footing with Both, implying that Text is not a supporting mode in this case but can stand by itself. One main advantage of Text over Voice is that it can be scanned randomly at whatever speed the user wishes, whereas Voice is a sequential output mode where the speed at which information can be acquired is not under the control of the user. The potential advantage of Voice over Text that users can scan other material visually while listening to Voice is not present in our experiments, since Voice was a duplication of Text. However, a number of users indicated that the combination of Both helped them to remember the information better. The output mode preferences of the non-experts were similar to the results shown for the experts.

*** Insert Figure 2 about here ***

*** Insert Table 2 about here ***

To further analyze the impact of output mode on user preferences, a two-way analysis of variance was performed on the scores that domain experts assigned to the four different apartments in each scene. Output mode and Apartment were used as factors, and the results were significant ($F=18.9, 4.97, 14.73, 21.56$ respectively, with $n_1=3$, $n_2=84$, $p < .01$) for the Apartment factor, for each of the four different scenes. On the other hand, neither factor interactions nor Mode were even marginally significant ($p > .10$) in any of the scenes, an indication that the output mode had little if any direct impact on user ratings for the apartments. Thus, any differences observed in efficiency or user preferences during the decision making process do not appear to have affected the decision outcomes.

5.3 Scene Completion Times: Efficiency

The SceneLength, (completion time) averaged over the 96 scenes evaluated by the domain experts (excluding scoring button times) was 307 seconds, with a standard deviation of 123 seconds. A two-way analysis of covariance was carried out for SceneLength, with Style (Analytic, Neutral, Heuristic) and Output Mode (Text, Voice, Both) as factors. Order, Gender and ButtonNo were used as concomitant variables. The design was unbalanced, requiring a least squares approach. The recommended techniques for analyzing unbalanced designs (Appelbaum & Cramer 1974) were followed throughout this study. In this paper, n_1 and n_2 are the degrees of freedom in the factor or concomitant variable and the error term respectively, while F is the partial F ratio for the measure in question. Interactions between the factors in this case were not significant ($p > .05$). However, main effects on both factors were significant; Style ($F=4.71$, $n_1=2$, $n_2=84$, $p<.05$) and Output Mode ($F=16.08$, $n_1=2$, $n_2=84$, $p<.01$). Plots of the main effects for both factors appear in Figures 3a and 3b respectively. A Tukey-Cramer statistical analysis (Neter, Wasserman & Kutner 1985) of paired comparisons of the results shown in each of these diagrams showed that all differences were significant ($p<.05$).

Of particular interest is the fact that Analytic individuals averaged approximately 18% longer per scene analysis than the Heuristic individuals, while Neutral individuals were between these two classes in completion times. This is in agreement with studies (Zmud 1978) which have shown that, in a situation unassisted by decision aids, Analytic individuals tend to prefer more information and to take more time to make a decision than do Heuristic individuals.

*** Insert Figure 3a about here ***

*** Insert Figure 3b about here ***

In this analysis the three concomitant variables were significant for the domain experts: ButtonNo ($F=432.7$, $n_1=1$, $n_2=84$, $p<.001$), Order ($F=7.15$, $n_1=1$, $n_2=84$, $p<.01$), and Gender ($F=4.28$, $n_1=1$, $n_2=84$, $p<.05$). For the 96 scenes analyzed, the average of SceneTotalNo (number of attribute references per scene) was 53.4, with a standard deviation of 20.75. On average, subjects took 26.5 seconds longer (9%) to complete the first scene task than the second scene, due to interface learning effects, and female subjects took an average of 20 seconds longer (7%) per scene than male subjects to complete their analyses.

In a SceneLength analysis similar to the above, with data gathered from domain non-experts, we found a mean of 338.9 seconds and a standard deviation of 128.1. However, only the main effect for Mode was significant ($F=4.95$, $n_1=2$, $n_2=36$, $p<.05$), as were the concomitant variables Order ($F=4.95$, $n_1=1$, $n_2=36$, $p<.05$) and Attribute References ($F=21.22$, $n_1=1$, $n_2=36$, $p<.01$). On average, non-expert subjects took 36 seconds longer to analyze the first apartment scene than the second (an approximate 10% time difference).

5.4 Effective Time Per Attribute Reference

A three way analysis of covariance was performed on the ButtonCutoffs for the domain experts, normalized to the number of buttons (1, 4 and 14 respectively) for the three abstraction levels. Level, Mode, and Style were used as the factors, and Order as a concomitant variable. This allowed an analysis of the time taken per attribute selected, as a percentage of the time the message was displayed (the "master" display time for that message). This highlights the time differences between voice and text modes. Three-way interactions were not modeled. None of the two-factor interactions were significant ($p>.05$ for each of the three pairwise interactions). The main effects for Level were not significant ($F=2.56$, $n_1=2$, $n_2=246$, $p>.10$), but main effects for Mode ($F=30.4$, $n_1=2$, $n_2=246$, $p<.001$), and Style ($F=8.62$, $n_1=2$, $n_2=246$, $p<.001$) were significant. The concomitant variable Order was also significant ($F=3.88$, $n_1=1$, $n_2=246$, $p<.05$), with the mean percentage cutoff being 7.5% less on the second scene than the first scene. This indicates that expert subjects were becoming more efficient in their information scanning processes by the time the first scene analysis had been completed.

The results of this analysis are displayed in Figures 4a and 4b, for the Mode and Style main effects, respectively. Differences among the main effects for Mode in Figure 4a are all significant ($\alpha=.05$). Differences among the main effects for Style are significant, except for the difference between Analytic and Neutral. The Order effect is shown by the two parallel lines for the corresponding main effect, with the higher line representing the results for the first scene task performed, and the lower line represents the second scene task. As expected, the Voice cutoff is at a higher percentage than Text, and Both is between these results. This suggests that the Voice and Text combination has the effect of lengthening the average time taken before the user goes on to access another attribute. We believe this indicates that some of the expert subjects were ignoring the text output in this case, listening to the voice information instead.

Expert Analytic subjects took longer than Heuristics, with the results for Neutral subject times lying between, in line with our previous observations on scene completion times.

*** Insert Figure 4a about here ***

*** Insert Figure 4b about here ***

A similar analysis of normalized ButtonCutoffs for the domain non-experts revealed that none of the two-way interactions were significant. The main effects for Mode ($F=21.8$, $n_1=2$, $n_2=117$, $p<.01$), Style ($F=4.72$, $n_1=2$, $n_2=117$, $p<.01$), and the concomitant variable Order ($F=13.8$, $n_1=1$, $n_2=117$, $p<.01$) were significant.

5.5 Analysis Of Average References Per Attribute

A three way analysis of variance was performed on ButtonNo for the domain experts, averaged according to abstraction level (the average number of references per attribute per scene at each level), with Level, Style, and Mode as the three factors. Three way interactions were not included in the analysis. For the domain expert data, we found that interactions between Mode and Level, and Mode and Style, were not significant. The main effects for Mode were also not significant ($F=2.11$, $n_1=2$, $n_2=247$, $p>.10$). However, there was a significant interaction between Level and Style ($F=2.83$, $n_1=3$, $n_2=247$, $p<.05$). These results appear in Figure 5. Using the Tukey-Cramer method of paired comparisons on these data, the differences in average references among the cognitive styles are not significant ($p>.05$) at any of the three levels. However, there is a significant increase in references from the General level to the Overall level for Neutral and Heuristic subjects. This difference is also significant for all three styles, when comparing the Overall level to the Specific level. An equivalent analysis of variance for the data from the non-experts gave no significant results for any of the main effects or interactions.

*** Insert Figure 5 about here ***

The average fraction of available attributes which were not referenced at particular levels by the domain experts during scene analysis was .12, .27, and .23 respectively, for the Overall, General, and Specific levels. This means that, for example, in the 96 scene analyses that were carried out by these subjects, an average of 23% of the Specific level attributes were not referenced at all, but conversely 77% were referenced at least once. This is considerably higher than the 41.3% reported by Payne (1976) in the closest comparable measurement he made, which had four alternatives and twelve information dimensions at what would correspond to our Specific abstraction level. The higher average references from our experiment is probably due to the fact that our system was computer-based, making access to data much easier than it was with Paynes's paper-based manual system. The corresponding average fractions of available attributes not referenced by non-experts was .11, .11 and .16 at the Overall, General, and Specific levels. A comparison of these fractions with those noted above for the domain experts is another indication that the non-experts were less selective and tended to gather information on a wider distribution of attributes, reinforcing the indication that they had a less structured view of the information needed in choosing an alternative.

The overall average number of references per attribute per scene at the three abstraction levels combined, was 2.80 for the domain experts. By abstraction level, the average number of references was 4.24, 2.40, and 2.84 per scene for attributes at the Overall, General, and Specific levels respectively. The corresponding average number of references by level for the non-experts was 3.70, 3.21 and 3.21 respectively. The latter result highlights the smaller factor differences in the analysis of variance by level for non-expert attribute references reported above.

6. Discussion

This research leads to several conclusions, relevant to multimodal information access and to the use of abstraction in information retrieval and display systems for decision makers.

1. There was no difference between domain experts and non-experts in their indicated preferences for output mode. Considering the total population of experimental subjects, there was no significant difference in preference between Voice and Text or between Both and Text.

However, Both was preferred to Voice. This indicates to us that, in combination, voice plays a supporting role to text rather than text playing a supporting role to voice. This is probably due to the fact that text is randomly scanned at the user's preferred speed, while voice is a sequential medium which proceeds at its own set speed.

2. Text as an information medium was approximately 30% faster than voice, as shown by our data analysis, and more convenient for the user although voice appeared to play a useful role by reinforcing text information (and perhaps assisting in remembering) information displayed in the Both mode. This improved efficiency did not appear to affect user preferences.

3. The ranking of alternatives and hence the decision made was not dependent upon the output mode, which results in more flexibility in the choice of output mode for decision situations such as this.

4. Cognitive style was useful in explaining several of the measures of domain expert performance. For example, the amount of time to complete a scene analysis was 18% less for heuristic than for analytic individuals, and the effective time per attribute reference was also about 15% less. There was also a small but significant difference between heuristic and analytic users in the frequency of access for abstract information (heuristic users tended to prefer abstract information more).

5. Several measures indicated that the lack of organization or structure in data gathering by the non-experts has an impact on the manner in which data is referenced. More attributes tended to be referenced by non-experts than by experts, and higher abstractions were not used as much. An entropy calculation confirmed that experts were significantly less random in their retrieval of information attributes.

We believe that experiments of this nature are helpful in pointing the direction for decision support interface development. Our results may be generalizable, and we plan future research to consider interfaces which use images in conjunction with other output modes. Based on the results obtained in this work, it appears that effective design for decision support that incorporates information abstraction, can be useful to domain experts if information of most value at lower levels is summarized at higher levels. User preferences and choice of output mode in our experiments, although they may impact efficiency, do not seem to affect the decisions made, provided that the same information is presented in each case. Finally, although some of the effects we observed were not large, these differences can loom large to a user who has preferences outside the parameters of an interface which must be used. It is therefore important to keep the broad user community in mind when such interfaces are developed, to provide usable multimodal systems for the user spectrum ranging from domain expert to non-expert, and across the spectrum of cognitive styles.

References

Anderson, J.R. 1985, *Cognitive Psychology And Its Implications* (New York: W.H. Freeman).

Appelbaum, Mark I. & Cramer, Elliot M. 1974, Some problems in the nonorthogonal analysis of variance, *Psychological Bulletin*, **81**, 335-343.

Archer, Norman P. & Kao, Diana. 1993, An empirical study of abstraction in conceptual model design, *ASAC Banff Conference (Information Systems) Proceedings*.

Baggett, P. & Ehrenfeucht, A. 1986, Encoding and retaining information in the visuals and verbals in an educational movie, *Educational Communications And Technology Journal*, **31** (3), 23-32.

Bariff, M.L. & Lusk, E.J. 1977, Cognitive and personality tests for the design of management information systems, *Management Science*, **23**, 820-829.

Batra, Dinesh, & Davis, Joseph G. 1989, A study of conceptual data modeling in database design: similarities and differences between expert and novice designers, in Janice I. DeGross, John C. Henderson, and Benn R. Konsynski (Eds.), *Proceedings, Tenth International Conference On Information Systems*, (New York: ACM), 91-99.

Benbasat, Izak, & Taylor, Ronald N. 1978, The impact of cognitive style on information system design, *MIS Quarterly*, **2**, 43-54.

Blaylock, Bruce K., & Rees, Loren P. 1984, Cognitive style and the usefulness of information, *Decision Sciences*, **15**, 75-91.

Bly, Sara A., Harrison, Steve R., & Irwin, Susan. 1993, Media spaces: Bringing people together in a video, audio, and computing environment, *Communications Of The ACM*, **36**, 28-47.

Bond, A.H., & Soetarman, B. 1988, Multiple abstraction in knowledge-based simulation, in Troy Henson (Ed.) *Artificial Intelligence And Simulation*, (San Diego, CA: Society for Computer Simulation), 61-66.

Chalfonte, Barbara L., Fish, Robert S., & Kraut, Robert E. 1991, Expressive richness: comparison of speech and text as media for revision, *CHI '91 Conference Proceedings*, (Reading, MA: Addison Wesley), 21-26.

Davis, Donald L., Barnes, James H. Jr., & Jackson, Wade M. 1993, Integrating communications theory, cognitive style and computer simulation as an aid to research on implementation of operations research, *Computers & Operations Research*, **20**, 215-225.

DeHaemer, Michael J. & Wallace, William A. 1992, The effects on decision task performance of computer synthetic voice output, *International Journal of Man-Machine Studies*, **36**, 65-80.

Guindon, R., Krasner, H., & Burtis, B. 1987, Cognitive processes in software design: activities in early, upstream design, in H.J. Bullinger and B. Shackel (Eds.) *Proceedings, Interact '87*, (Amsterdam: Elsevier), 383-388.

Hodges, Matthew E., & Sasnett, Russell M. 1993, *Multimedia Computing: Case Studies From Project Athena*, Reading, MA: Addison Wesley.

Huber, George P. 1983, Cognitive style as a basis for MIS and DSS designs: Much ado about nothing? *Management Science*, **29**, 567-579.

Miller, G.A. 1956, The magic number seven plus or minus two: Some limitations on our capacity to process information, *Psychological Review*, **63**, 81-97.

Myers, Isabel Briggs, & McCaulley, Mary H. 1985, *Manual: A Guide To The Development And Use Of The Myers-Briggs Type Indicator*, (Palo Alto, CA: Consulting Psychologists Inc.).

Neter, John, Wasserman, William, & Kutner, Michael H. 1985, *Applied Linear Statistical Models*, (Second Ed.), (Homewood IL: Irwin).

Nielsen, Jakob. 1993, *Usability Engineering*, (Boston, MA: Academic Press).

Nugent, G. 1982, Pictures audio and print: symbolic representation and effect on learning, *Educational Communication And Technology Journal*, **30** (3), 163-174.

O'Keefe, Robert M., & Pitt, Ivan L. 1991. Interaction with a visual interactive simulation, and the effect of cognitive style, *European Journal Of Operational Research*, **54**, 339-348.

Ossher, H.L. 1987, A mechanism for specifying the structure of large, layered systems, in Bruce Shriver and Peter Wegner (eds.) *Research Directions In Object-Oriented Programming*, (Cambridge, MA: MIT Press), 219-252.

Payne, J.W. 1976, Task complexity and contingent processing in decision making: A replication and protocol analysis, *Organizational Behavior And Human Performance*, **16**, 366-387.

Ramaprasad, A., & Mitroff, I.I. 1984, On formulating strategic problems, *Academy of Management Review*, **9**, 597-605.

Shannon, C.E. 1948, A mathematical theory of communication, *Bell System Technical Journal*, **27**, 379-423 & 623-656.

Staggers, Nancy, & Norcio, A.F. 1993, Mental models: Concepts for human-computer interaction research, *International Journal of Man-Machine Studies*, **38**, 587-605.

Streeter, Lynn A. 1990, Applying speech synthesis to user interfaces, in M. Helander (ed.) *Handbook Of Human-Computer Interaction*, (Amsterdam: Elsevier), 321-343.

Taggart, William, & Robey, Daniel. 1981, Minds and managers: On the dual nature of human information processing and management, *Academy of Management Review*, **6**, 187-195.

Todd, Peter, & Benbasat, Izak. 1993, An experimental investigation of the relationship between decision makers, decision aids and decision making effort, *INFOR*, **31**, 80-100.

Zmud, R.W. 1978, Individual differences and MIS success: A review of the empirical literature, *Management Science*, **25**, 966-979.

Appendix

Data For One Apartment Choice

Attribute[1]	Data
Overall	One bedroom basement apartment, rather low priced, good condition.
Environment	Moderately quiet and well maintained.
Landlord	Visits frequently.
Noise	Near an elementary school.
Cleanliness	Virtually spotless.
Brightness	Dark, with one medium and two small windows.
Rental	Rather low priced with a long term lease.
Rate	$265 per month plus $65 for utilities.
Lease	Twelve month lease.
Location	Downtown Hamilton.
Campus	Twenty minute drive to campus.
Shopping	Large shopping centre is two blocks away.
Bus	Stop for buses going to campus is one block away.
Parking	Outdoor parking.
Interior	Small one bedroom apartment; living room, kitchen, full bath, two closets.
Size	8 x 10 foot bedroom, 12 x 12 foot living room.
Closet	One closet in the bedroom, and one in the hallway.
Kitchen	Dishwasher and dishes available.
Features	Coin operated laundry available.

[1] Amount of indentation is related to abstraction level, with "Overall" at the highest level.

List Of Tables And Figures

Table 1

Attribute Importance Ranking

	Rating Importance Rank		Mean Number of References		Rank by Mean Number of Refs.	
Attribute	**Expert**	**Non-Expt**	**Expert**	**Non-Expt**	**Expert**	**Non-Expt**
Overall**	10	9	4.25	3.70	2	4
Environ*	16	12	2.32	3.09	13	10
Landlord	9	11	3.40	3.91	5	3
Noise	12	7	3.00	3.34	9	8
Clean	3	4	3.46	3.70	4	5
Bright	6	18	3.05	2.25	8	18
Rental*	11	5	1.90	3.39	16	6
Rate	1	1	6.04	5.84	1	1
Lease	2	8	3.24	3.18	6	9
Location*	4	2	3.15	3.36	7	7
Campus	5	3	3.47	3.98	3	2
Shopping	18	15	1.27	2.70	18	16
Bus	8	17	2.60	2.73	11	15
Parking	17	14	1.69	2.64	17	17
Interior*	14	6	2.19	3.00	15	13
Size	7	10	2.50	3.05	12	12
Closet	19	19	0.96	1.66	19	19
Kitchen	13	13	2.29	2.93	14	14
Features	15	16	2.66	3.05	10	11

Table 2

Interface Comparison Preference Results

Question	Voice vs. Text				Voice vs. Both				Text vs. Both			
	n	p	sig.	Pref.	n	p	sig.	Pref.	n	p	sig.	Pref.
1	13	.29	ns	-	15	.000	***	Both	12	.61	ns	-
2	16	.40	ns	-	16	.000	***	Both	14	.91	ns	-
3	14	.79	ns	-	16	.000	***	Both	12	.93	ns	-
4	16	.40	ns	-	14	.001	***	Both	13	.011	*	Text
5	14	.21	ns	-	14	.001	***	Both	15	.30	ns	-
6	12	.39	ns	-	16	.000	***	Both	15	.15	ns	-
7	8	.64	ns	-	16	.000	***	Both	12	.61	ns	-
8	16	.60	ns	-	16	.000	***	Both	15	.70	ns	-

Notes:

a) Domain expert data collected from preference questionnaire (Figure 2)

b) Sign test used for comparisons (tied rankings discarded, resulting in differing n values)

c) n = sample size, p = sig. level, sig. = ns (not sig.), * (.05 level), *** (.001 level), Pref. = significant preference of the output mode pair tested.

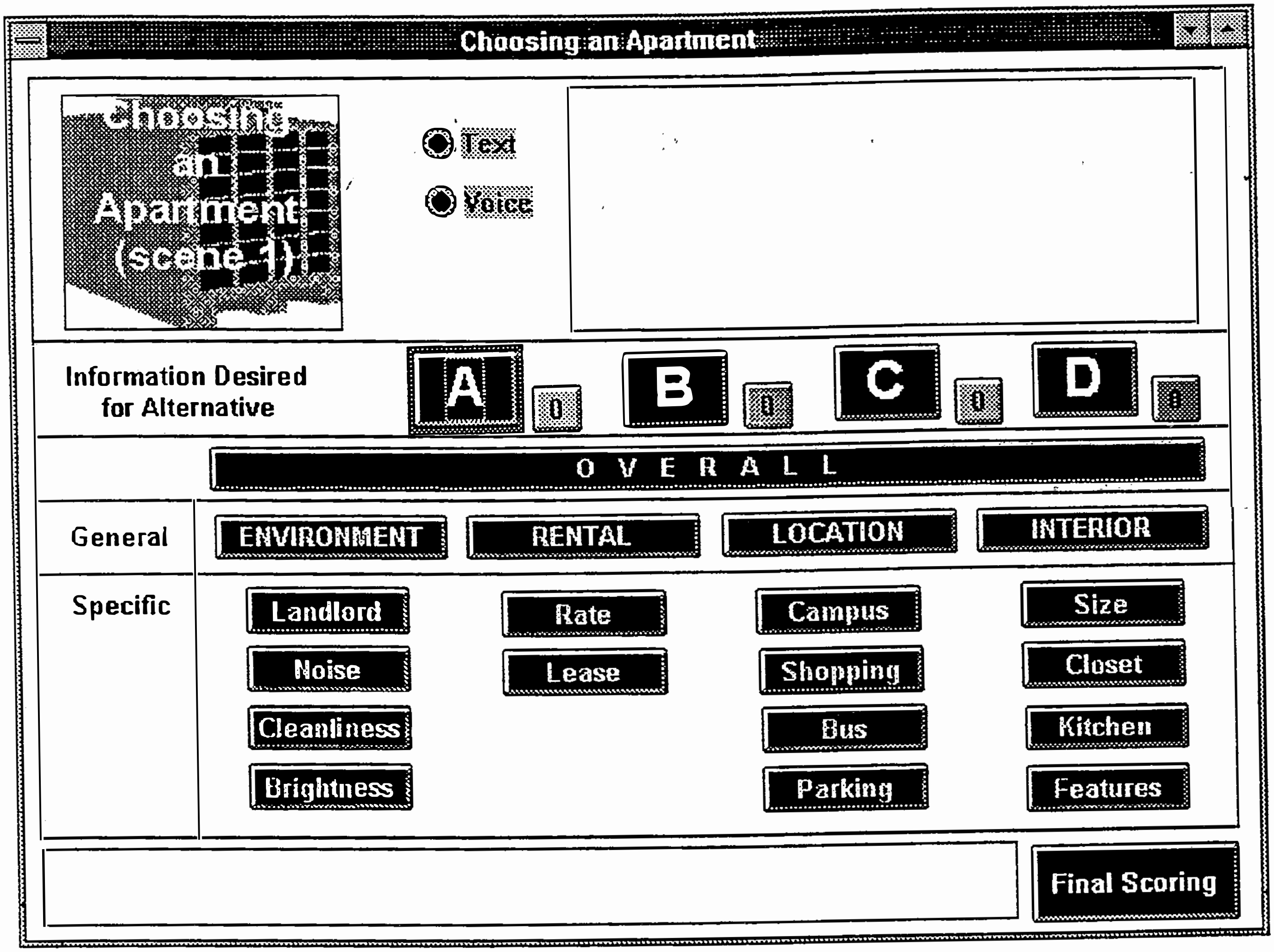
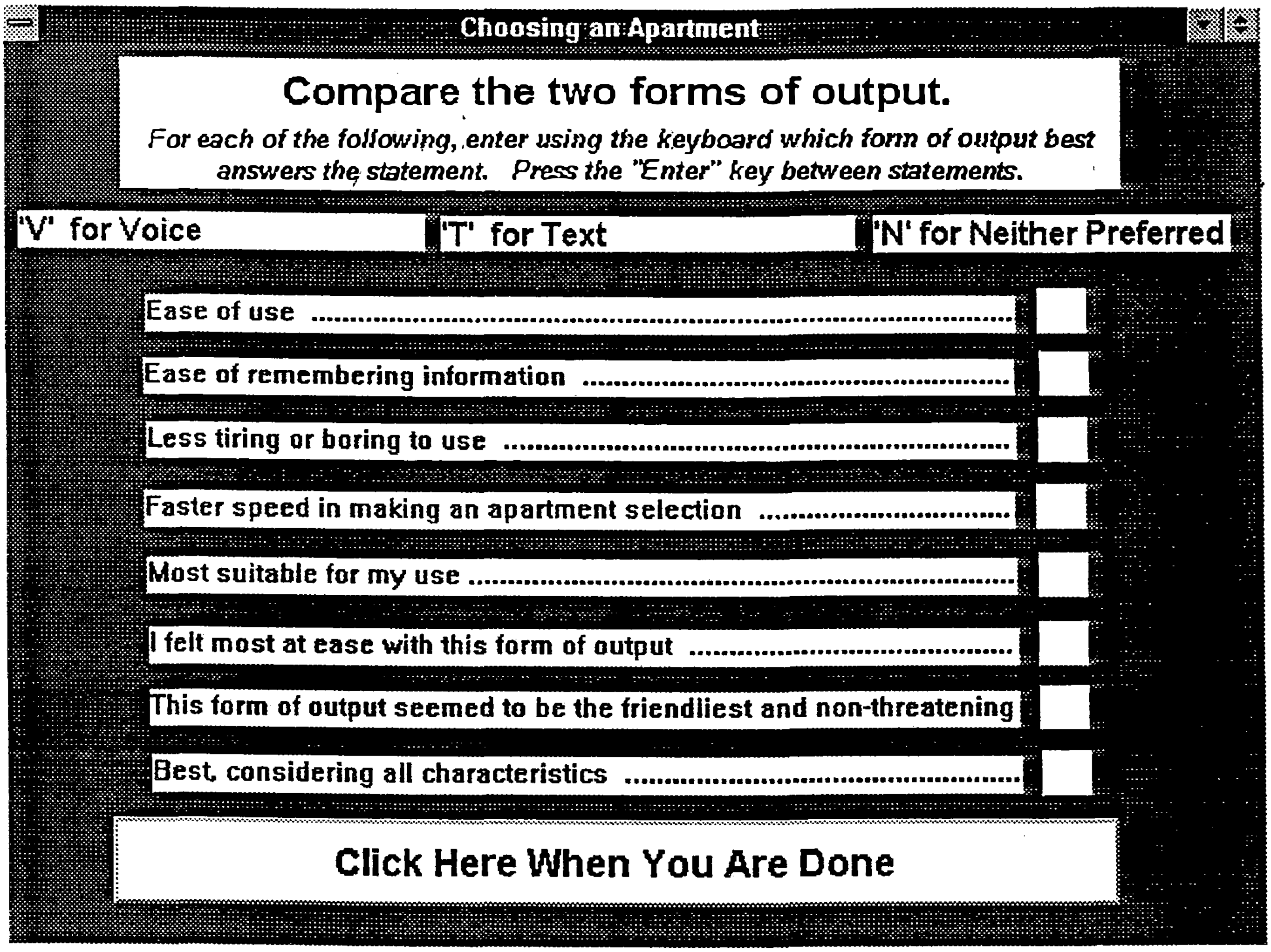

Choosing an Apartment
Choosing an Apartment (scene 1)
Text
Voice
Information Desired for Alternative
A
0
B
0
C
0
D
0
O V E R A L L
General
ENVIRONMENT
RENTAL
LOCATION
INTERIOR
Specific
Landlord
Noise
Cleanliness
Brightness
Rate
Lease
Campus
Shopping
Bus
Parking
Size
Closet
Kitchen
Features
Final Scoring

FIGURE 1



Choosing an Apartment

Compare the two forms of output.

For each of the following, enter using the keyboard which form of output best answers the statement. Press the "Enter" key between statements.

'V' for Voice | 'T' for Text | 'N' for Neither Preferred

Ease of use ....

Ease of remembering information ....

Less tiring or boring to use ....

Faster speed in making an apartment selection ....

Most suitable for my use ....

I felt most at ease with this form of output ....

This form of output seemed to be the friendliest and non-threatening

Best, considering all characteristics ....

Click Here When You Are Done

FIGURE 2

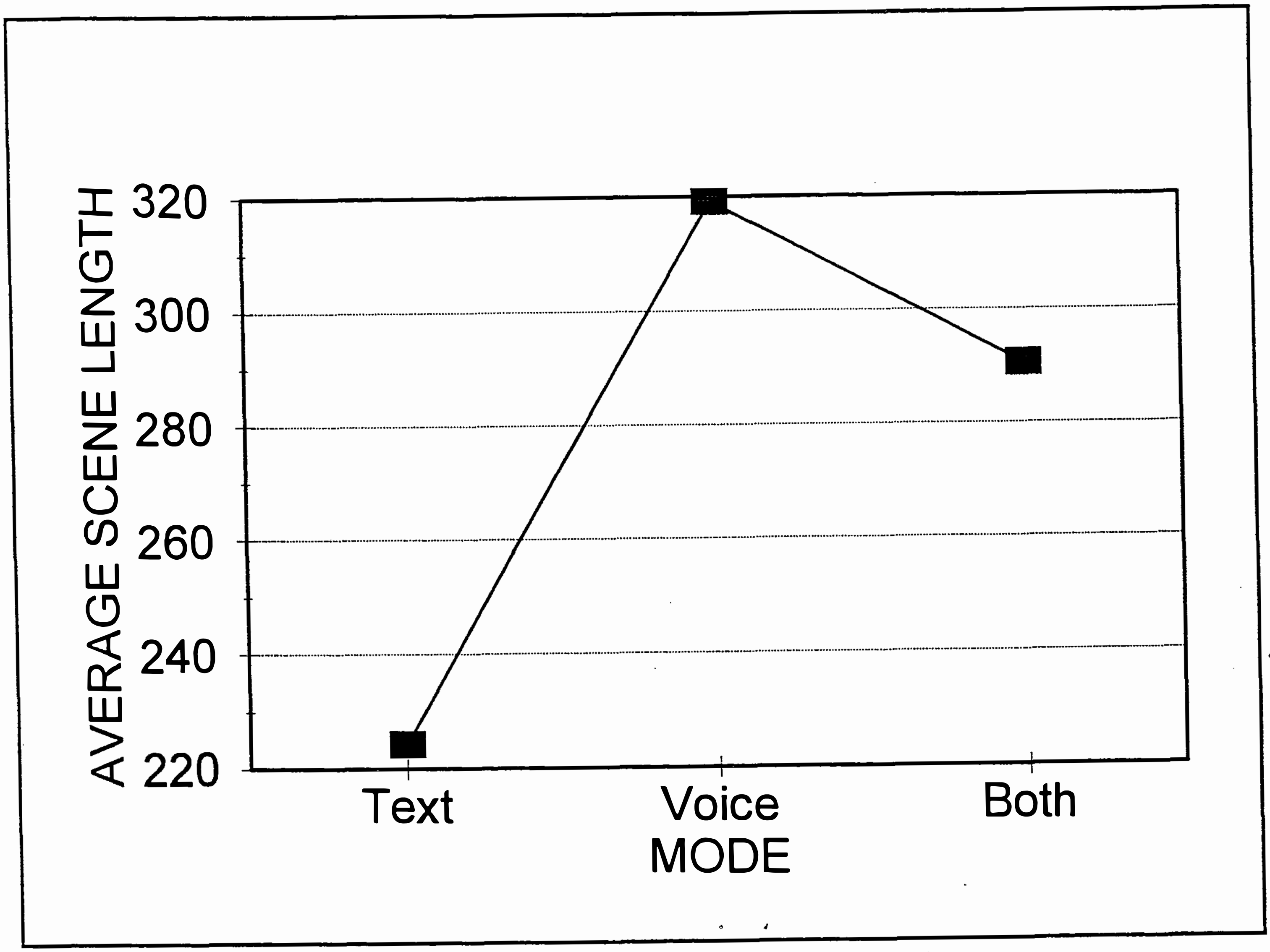

AVERAGE SCENE LENGTH
320
300
280
260
240
220
Text
Voice
Both
MODE

FIGURE. 3a

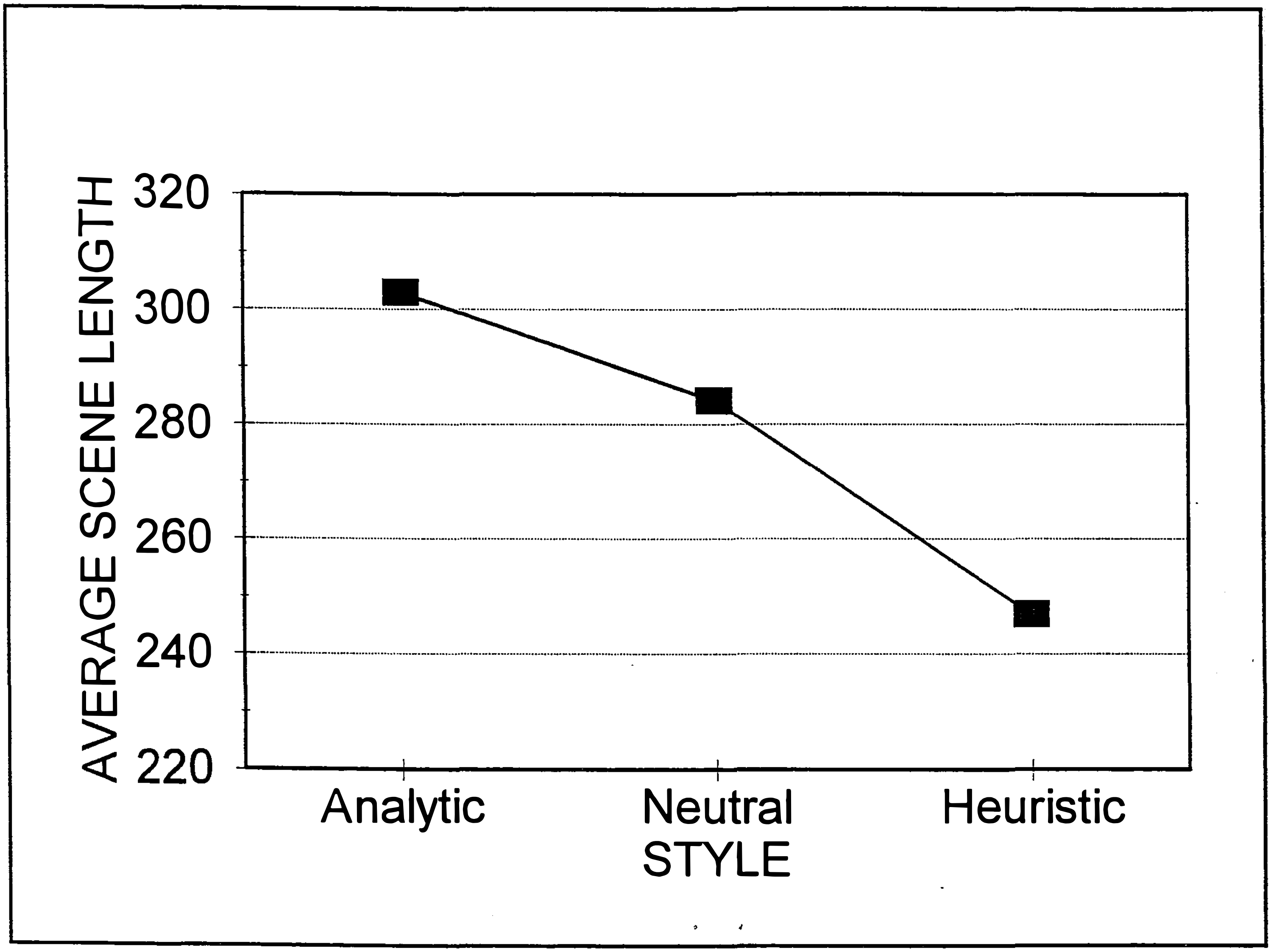
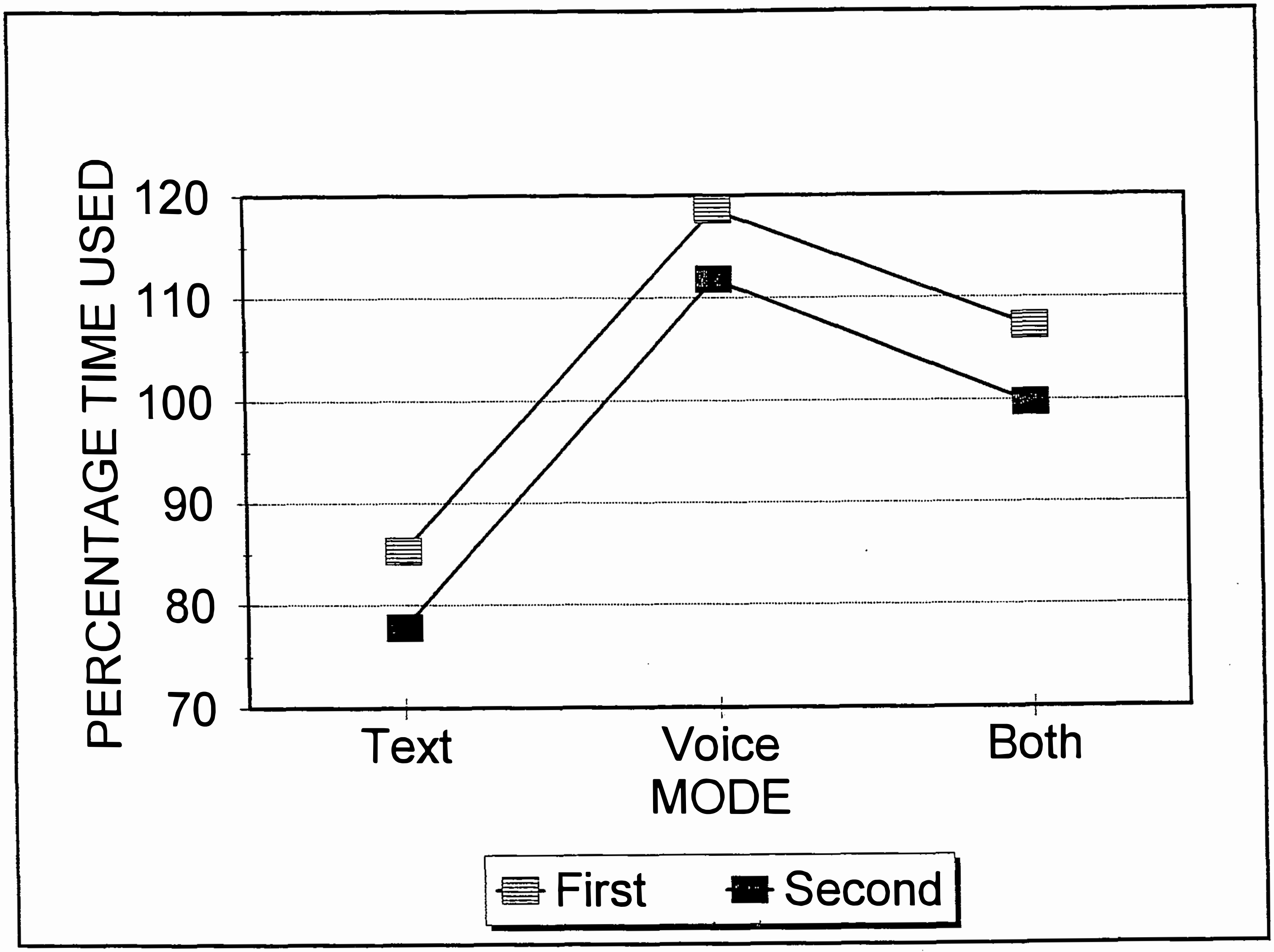
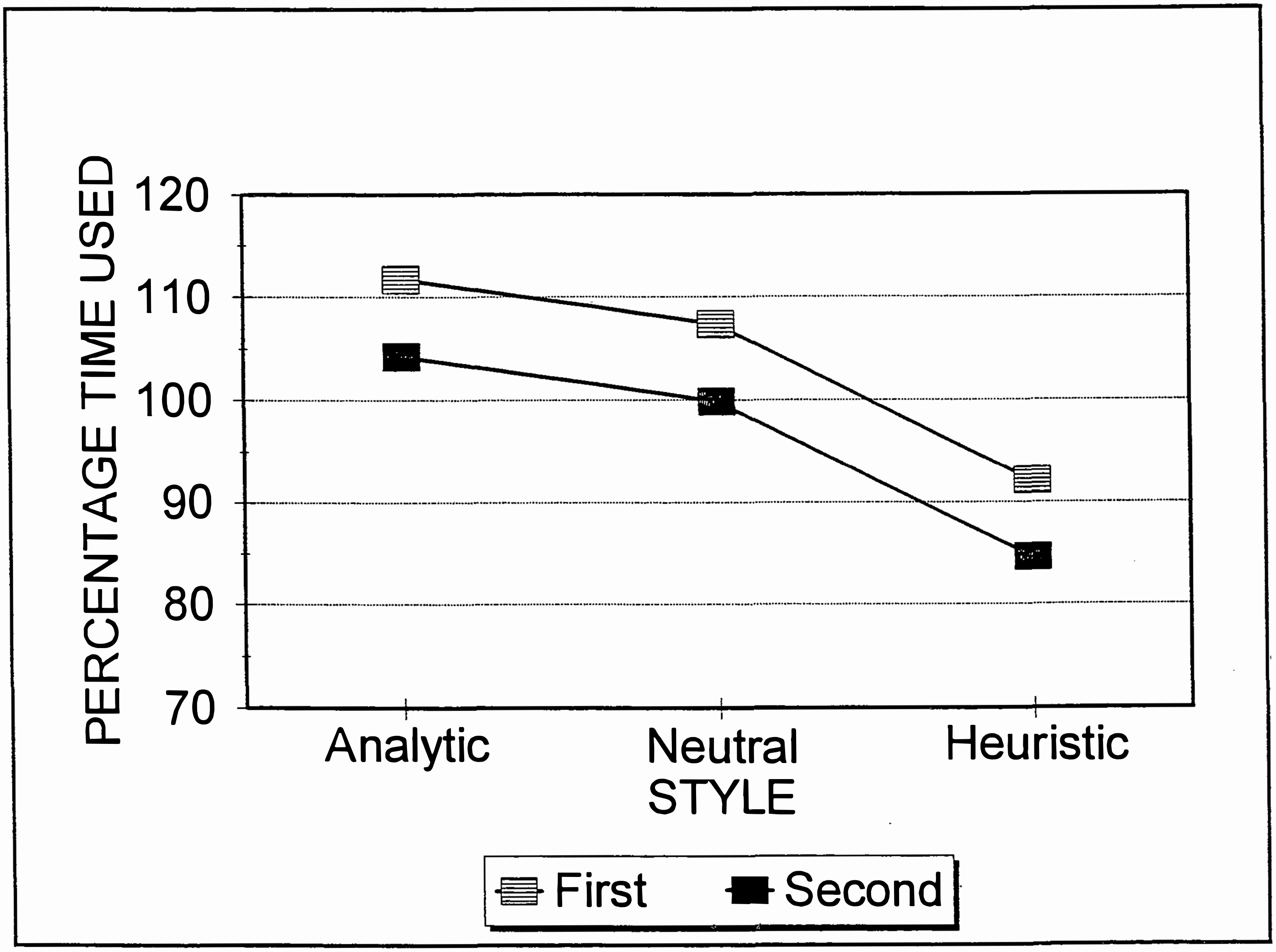

AVERAGE SCENE LENGTH
320
300
280
260
240
220
Analytic
Neutral
Heuristic
STYLE

FIGURE 3b



PERCENTAGE TIME USED
120
110
100
90
80
70
Text
Voice
Both
MODE
First
Second

FIGURE 4a



PERCENTAGE TIME USED
120
110
100
90
80
70
Analytic
Neutral
Heuristic
STYLE
First
Second

FIGURE 4b

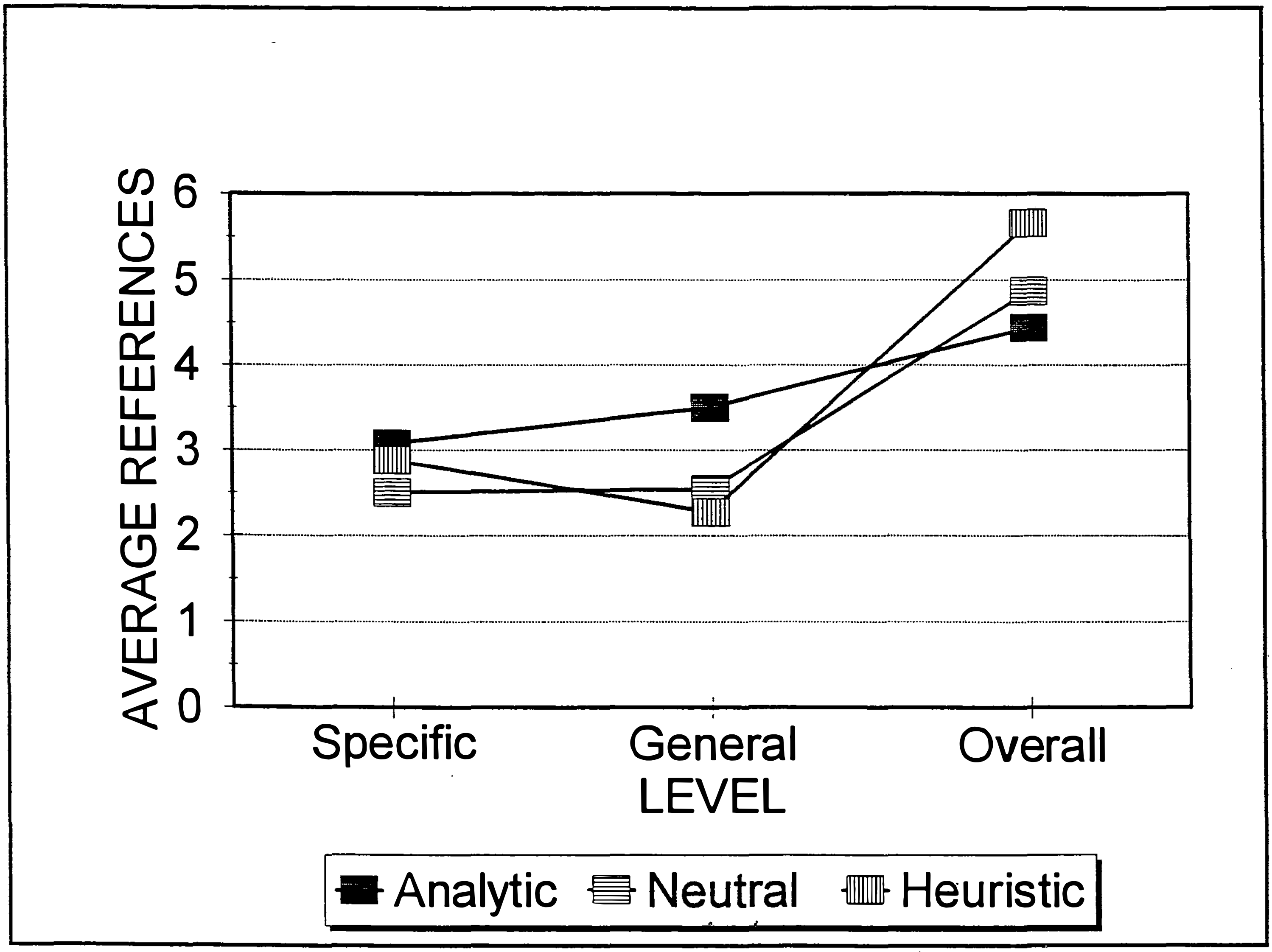

AVERAGE REFERENCES
0
1
2
3
4
5
6
Specific
General
Overall
LEVEL
Analytic
Neutral
Heuristic

FIG. 5.

School of Business
McMaster University

WORKING PAPERS - RECENT RELEASES

361. Roy J. Adams, "Employment Relations in an Era of Lean Production", April, 1991.

362. John W. Medcof, John Bachynski, "Measuring Opportunities to Satisfy the Needs for Achievement, Power and Affiliation", April, 1991.

363. Robert F. Love, Paul D. Dowling, "A Nested Hull Heuristic for Symmetric Traveling Salesman Problems", June, 1991.

364. Robert F. Love, John H. Walker, and Moti L. Tiku, "Confidence Levels for $\ell_{k,p,\theta}$ Distances", June, 1991.

365. Archer, N. P., and G. O. Wesolowsky, "A Dynamic Service Quality Cost Model with Word-of-Mouth Advertising", July, 1991.

366. John W. Medcof, "Weiner, Kelley, Jones, and Davis, Peat: Integrated", August, 1991.

367. Harish C. Jain and Rick D. Hackett, "A Comparison of Employment Equity and Non-Employment Equity Organizations On Designated Group Representation And Views Towards Staffing", August, 1991.

368. J. Brimberg and R. F. Love, "Local Convergence in a Generalized Fermat-Weber Problem", August, 1991.

369. Rick D. Hackett and Peter Bycio, "The Temporal Dynamics Associated with Absenteeism as a Coping Mechanism", September, 1991.

370. Robert F. Love and John H. Walker, "A New Criterion For Evaluating the Accuracy of Distance Predicting Functions", October, 1991.

371. Joseph B. Rose, "Interest and Rights Disputes Resolution in Canada, New Zealand and Australia", October, 1991.

372. Paul A. Dion and Peter M. Banting, "Buyer Reactions to Product Stockouts in Business to Business Markets", October, 1991.

373. Isik U. Zeytinoglu, "Part-Time and Other Non-Standard Forms of Employment: Why are They Considered Appropriate for Women?", November, 1991.

374. Rick D. Hackett, Peter Bycio and Peter Hausdorf, "Further Assessments of a Three-Component Model of Organizational Commitment", January, 1992.

375. Y. Lilian Chan, Bernadette E. Lynn, "Evaluating Tangibles and Intangibles of Capital Investments", February, 1992.

376. Christopher K. Bart, "Strategic Lessons for Today's Airline CEO's", February, 1992.

377. Yufei Yuan, "A Stable Residence Exchange Problem", March, 1992.

378. Jim Gaa,, "The Auditor's Role: The Philosophy and Psychology of Independence and Objectivity". April, 1992.

379. David J. Buchanan and George O. Wesolowsky, "Algorithm for Rectilinear Distance Minimax Single Facility Location in the Presence of Barriers to Travel", April, 1992.

380. Thomas E. Muller and D. Wayne Taylor, "Eco-Literacy Among Consumers: How Much Do They Know About Saving Their Planet?", April, 1992.

381. Willi H. Wiesner, "The Contributions of Job Relevance, Timing, and Rating Scale to the Validity of the Structured Employment Interview", September, 1992.

382. Robert G. Cooper, "Third Generation New Product Processes", November, 1992.

383. Robert F. Love and John H. Walker, "Distance Data for Eighteen Geographic Regions", August, 1993.

384. Jack Brimberg and Robert F. Love, "New Hull Properties for Location Problems", August, 1993

385. Harish C. Jain and S. Muthuchidambaram, "Bill 40 Amendments to Ontario Labour Relations Act: An Overview and Evlauation", September, 1993.

386. Norm P. Archer and Randy Bassett, "Self-Organizing Feature Maps: The Traveling Salesman Problem and Other Applications", September, 1993.

387. John W. Medcof, "Need Satisfaction and Job Behaviours in Four Occupations", October, 1993.

388. Harish C. Jain and C.S. Venkata Ratnam, "Affirmative Action in Employment for the Scheduled Castes and the Scheduled Tribes in India", December, 1993.

389. Harish C. Jain and S. Muthu, "A Legislative Framework Restricting the Use of Replacement Workers: The Ontario Approach and Its Relevance to U.S. Labor Law Reform", December, 1993.