Interacting with Computers vol 8 no 4 (1996) 323–345

Investigation of voice and text output modes with abstraction in a computer interface

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

A human–computer interface is described, which was designed to study user preferences and the effectiveness of output modes and levels of information abstraction in a decision making environment. The interface was tested in an exploratory study of an apartment selection problem. It was observed that text plus voice was preferred over voice alone, but there was no significant difference in preferences between text and voice or between text and text plus voice. This indicates that adding text to voice output improves the perceived acceptability of voice, but adding voice to text does not alter the perceived acceptability of text. The text mode was most efficient in performing information search, followed by voice mode and text plus voice mode in that order. We observed inconsistencies between the users' perceived importance of information attributes and the actual usage of these attributes, and inconsistencies between the perceived importance of and the usage of abstraction levels. We did not observe significant differences between users with task domain experience and those which did not have domain experience, but cognitive style did affect task performance. Our findings suggest that a user interface should either provide flexible access at different abstraction levels, or should organize information based on its perceived importance to the user rather than its level of abstraction.

Keywords: interfaces, output modes

Recent advances in multimedia computer equipment and software Bly *et al.*, 1993; Hodges and 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. Previous studies into how such systems can most effectively support the decision process or how they may impact decision outcomes have been limited primarily to comparisons of graphical and tabular data presentation modes (summarized in a meta-analysis by Montazemi and

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

0953-5438/96/$09.50

Wang, 1989). Vessy (1991) showed that the differences in user performance in these cases were due to differences in the matching of task type (spatial *versus* symbolic) with data presentation. That is, a graphical representation is better suited to a spatial task, and a tabular presentation is better suited to a symbolic task.

Multimedia computing can add many other dimensions to human–computer communication, 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.). The challenge is to design effective interfaces which take advantage of these additional communication modes. In addition to the presentation mode, information can also be ordered differently or in differing levels of detail. The designer may also need to consider multiple objectives, which can include minimizing information accessing effort, maximizing communication efficiency, matching user preference to presentation mode, and providing information in such a way 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, but research is needed to determine how balance can be achieved in this variety of decision-making situations.

Traditionally, text is the major input/output mode for human–computer communication, but voice communication between human and computer is now feasible through speech synthesis and voice recognition (Streeter, 1988). Since oral and written communication are the most widely used mode for verbal information exchange, it is natural to expect that people should be able to converse with machines as effortlessly as they converse with one another over the telephone. However, voice input/output technology is still not mature and requires more computing resources, including storage space and processing power, than text does. In this paper, we limit our study to the comparison of voice and text output modes only. To evaluate the most effective use of text or voice output, we investigate how either mode or their combination can affect user preferences and task performance in a decision-making situation. In addition, we examine the effects of information abstraction support at different levels, and the impact of certain individual differences on performance and user preference. In short, the objective of this research is to determine how text and voice output modes can be combined with different abstraction levels in interface design to assist users in acquiring decision information across a range of user characteristics and preferences. This type of study is essential to help point the way to more usable interface designs for the more powerful multimedia systems now available.

In the following, we first define the dimensions of the study and develop a set of propositions based on a review of previous work. We then describe the experiment used to test our propositions. Finally we analyse the experimental results and derive our conclusions.

Dimensions of the study

Two dimensions are under consideration: interface characteristics and user

characteristics. Interface characteristics are represented by the output mode (text or voice) and the levels of information abstraction. User characteristics are represented by user experience level and cognitive style.

Interface characteristics

We focus on two aspects of the interface: output mode (text or voice) and levels of information abstraction among attributes. With the current wide availability of multimedia computing, voice is often used in conjunction with visual information because it provides a complementary channel to assist users in assimilating information, information can be received without direct attention to the source, and the user can simultaneously attend to other tasks (Streeter 1988). Voice and text combinations provide a useful experimental platform because exactly the same information can be provided through each channel, giving an effective means of comparison. There are also a number of ways in which information can be represented at different levels of abstraction, in order to respond to user needs in assimilating information and reasoning about decisions. Output modes and information abstraction are further discussed below.

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 (1988) outlined the advantages and disadvantages of speech in comparison with text. The major advantage of using speech in an interface is its universality; almost everyone understands spoken language. But one notable disadvantage is that voice delivers information at less than half the rate that text can be scanned usefully. Any combination of voice and text is also likely to slow the information acquisition process. However, 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. Sipior and Garrity (1992) found that presentations with a mix of audio and visual accompaniments improved receptiveness attributes such as perception, attention, comprehension, and retention. 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.

Most previous studies either make intuitive comparisons between voice and text, based on their own characteristics, or compare them when they are used to represent different information or in different contexts. In our study, we single out the effects of the voice and text modes and make more precise performance

measurements. Given the indications of improved performance with combined voice and text, we anticipate that:

- *Proposition 1*: When representing the same information, a voice and text combination will be preferred by users over either voice only or text only, and text will be preferred over voice only.

It is clear that information is acquired more rapidly by scanning text rather than listening to voice. We believe that this will not be the case when text is combined with voice, and that voice output will tend to slow information acquisition and hence decision making. This leads to:

- *Proposition 2*: Information access will be faster with an interface that uses text output, compared to one with voice output or with a combination of voice and text output.

When the same information is made available through either text or voice output modes, it is anticipated that:

- *Proposition 3*: Decisions will not be affected by the voice or text output mode used in the interface when these modes contain the same information.

Abstraction levels

In a problem solving or reasoning process, it is useful to use information abstraction to reduce information overload, since people have limited short-term memory which can handle only about seven 'chunks' of information (Miller, 1956). At different stages of the information searching and evaluation process, a decision maker may need information to be represented at different levels of abstraction, from the higher levels containing less information of a more generalized nature, to the lower levels containing more detailed information of a more specific nature. 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 unnecessary details. Information abstraction finds application in diverse areas such as solving problems (Anderson, 1985), formulating strategic problems (Ramaprasad and Mitroff, 1984), enhancing creativity in problem solving (Couger *et al.*, 1993), systems design (Guindon *et al.*, 1987), and simulation modelling (Bond and Soetarman, 1988).

Levels of abstraction are often predetermined by system designers and are organized in a top-down hierarchical structure. Among the most commonly used techniques are menu selection and the use of windows in a user interface (Norman *et al.*, 1986). With these interfaces, users are forced to access information in a top-down manner. To assess the true preference and usage of information abstraction in interface design, it is important to ensure that users do not need to use different amounts of effort to access information at different abstraction levels, as Todd and Benbasat (1993) have shown that decision makers tend to access information in a manner which minimizes effort. In our experimental interface

design, we attempted to minimize any difference in effort required to access information at different abstraction levels. As the use of abstraction may depend on user characteristics, we identify the most important of these characteristics in the following section.

User characteristics

There are a number of individual characteristics which may have the potential to affect user reaction to different information presentation and arrangement styles. The most frequently used measures in experimental interface studies include user experience level and cognitive style. These are discussed below.

User experience. Nielsen (1993) indicates that user 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 people use data presented through a computer interface. For example, Archer and Kao (1993) found that users with domain experience were more likely to make use of high level abstractions in problem solving. Batra and Davis (1989) also found that users experienced in the task domain focused on generating an holistic understanding of a problem before solving it, 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, as this provides a framework or schema with which to structure known information (Staggers and Norcio, 1993). In our work, we chose subjects from populations which had similar computer system experience levels, but with two levels of experience in the task domain. Previous research leads to an expectation that:

- *Proposition 4*: Given the freedom to access information directly at different abstraction levels, experienced users will access higher level information abstractions more frequently than will inexperienced users.

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). However, as Umanath *et al.* (1990) point out, although only about 10% of variance in decision maker performance or behaviour can be attributable to cognitive style, research models in behavioural fields like MIS are unable to explain substantial proportions of variation in response variables. We cannot afford to ignore a variable that can account for 10% of the explainable variance.

One category of cognitive style is 'thinking mode' (Bariff and Lusk, 1977). Taggart and Robey (1981) suggested that a way to measure thinking mode for decision makers was to use two of the scales from the Myers-Briggs Type Indicators (MBTI) (Myers and McCaulley, 1985) instrument. These scales are:

- the Thinking/Feeling measure which contrasts rational judgements by

objective and logical analysis (Thinking) with weighing the relative person-centered values (Feeling);
- the Sending/Intuition measure of an interest in objects, where events and details of the present moment (Sensing) are contrasted with the possibilities, abstractions, and insights imagined for the future (Intuition).

On these two scales, there are four different outcomes for individual scores that represent dominant scores in either direction on each of the two scales, with possible combinations of ST (Sensing–Thinking), NT (Intuition–Thinking), NF (Intuition–Feeling), and SF (Sensing–Feeling). Taggart and Robey classified decision makers with ST scores as having an analytic decision-making style, and those with NF scores as having an heuristic decision-making style. The NT and SF types were 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. In related works, O'Keefe and Pitt (1991) found weak evidence that preference for display type can be partially explained by cognitive style. Davis *et al.* (1992) suggested that the method of communicating information to users will be more effective if it is matched to their cognitive style, and Blaylock and Rees (1984) concluded that cognitive style influences a decision maker's evaluation of an unstructured, strategic planning problem. Zmud (1978) found 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. Based on the foregoing, and in the context of the experiments we performed, we would expect that:

- *Proposition 5*: Analytic individuals will take more time to make decisions than will heuristic individuals.

The experiment

Experimental interface and decision tasks

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 computer interface was developed to support a simple decision-making task. The decision task was similar to, but implemented differently from, the apartment selection task introduced by Payne (1976). In Payne's original experiment, attribute values of apartments were printed on cards and shown on an information board. Todd and Benbasat (1993) and other researchers have used this experiment in a modified computerized form. Their research focus was on different decision-making strategies, rather than on the output modes and levels of abstraction. In our experiment, the task was set up so that each alternative selection had the same set of attributes in which different subjects might have different interests, when looking at such an apartment. Information attributes were both qualitative and quantitative, of the type normally used in making apartment choices. There was a total of 19 attributes including 14 attributes at the Specific (most detailed) level of abstraction, four

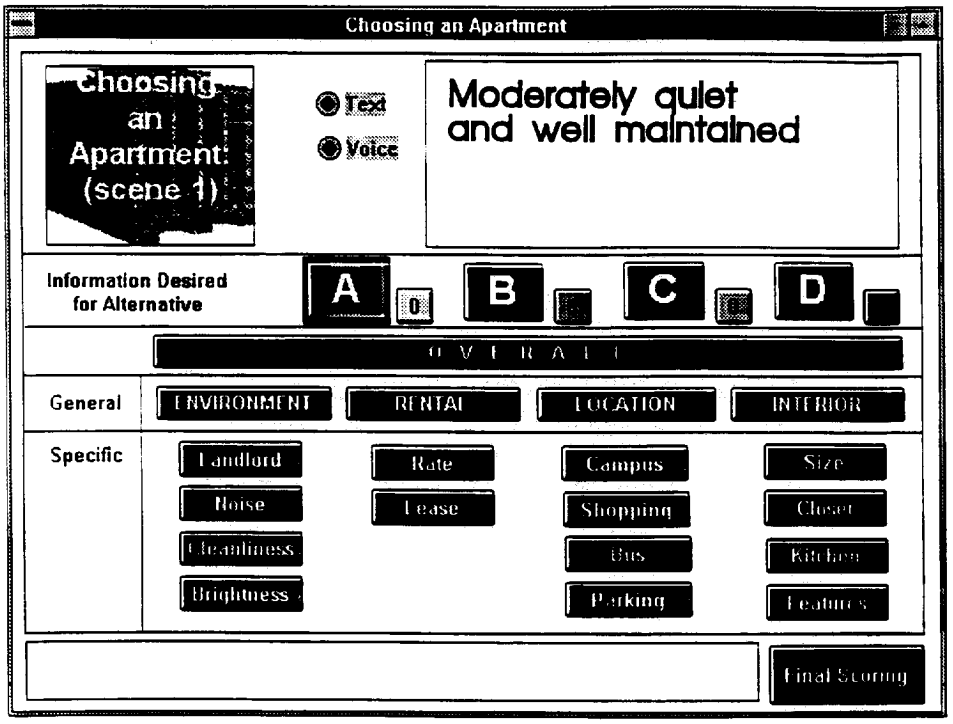


Figure 1. Apartment ranking interface

attributes at the General (intermediate) level of abstraction, and one attribute 'Overall' at the highest level of abstraction. Appendix 1 shows the attributes and their values for one of the apartments used in the study. The information was presented in small chunks in each attribute. The sequential nature of voice makes it necessary to present voice information in small and manageable chunks in order to reduce the disadvantages that it otherwise has in relation to text, which can be accessed randomly and at a viewer-controlled pace.

The interface displayed to users is shown in Figure 1. As this display gives equal access to all information attributes in a particular apartment, it avoids hampering access to information that could result in a bias towards selecting attributes at particular abstraction levels. This type of bias may occur in a hierarchical interface such as a menu system. Information was not displayed unless requested by the subject, and then only for a limited time. This arrangement is basically a shallow two-level menu, with the upper level allowing the selection of an apartment for consideration and the lower level allowing the selection of any apartment attribute for display.

As 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, which are referred to as 'scenes'.

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. The use of recorded voice rather than synthesized voice avoided the potential problems associated with low quality synthesized voice. Subjects used a mouse to select screen buttons at will to get information on particular attributes of the apartment in question, basically simulating a database browsing interface. The system recorded and time stamped each button press on a text file. These files and computerized questionnaire responses were used later for statistical analysis.

A scoring button shown beside each of the four apartments 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 rating for that 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 grey 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.

Three output modes could be selected in advance by the experimenter for each scene. These modes allowed output either entirely by voice only (Voice), by text only (Text), or by both voice and text output of identical information (Both). 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. However, the user could terminate an output and go on to another attribute at any time simply by clicking on the next attribute of interest. Text information was shown in a small pane at the upper right of the screen (see Figure 1). When voice was used, it duplicated the same information as the text output. 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.

This experiment differs from previous experiments of this nature (Payne, 1976; Todd and 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. This allowed precise measurements of what attributes were accessed, how often they were accessed, and when and for how long this occurred. Previous reported experiments of this type have usually displayed information continuously after it was accessed, requiring verbal protocol analysis to determine which attributes were consulted, thus introducing additional uncertainty into the measurement process.

Dependent variables

We measured both user preference and task performance for each interface session. User preference measures were collected from subjects through

computerized questionnaires. These gave the rank-ordered preferences for both output modes and information attributes. Task performance measures were collected automatically by recording button presses, and included the frequency of attribute selection and the total time spent observing or listening to information on each attribute during each scene. The ToolBook software enabled us to collect very detailed data during the experiment. The list of the dependent variables follows:

- **Mode preference:** Data were gathered after subjects completed their decision making tasks, through a questionnaire (Figure 2), where the subjects indicated a preference or indifference for either of the two output modes they used.
- **Attribute preference:** Data were gathered through another questionnaire where subjects indicated a preference rank ordering among all the information attributes, in terms of their perceived importance.
- **Attribute selection frequency:** Data was gathered during the experiment by measuring the total number of times a particular button was pressed (thus accessing that particular information attribute) during a sense. This gives an indication of the relative usage of that attribute by the subject.
- **Scene length:** This is the total time used by a subject on information gathering during a scene, and was used to evaluate the time efficiency of a subject in performing an apartment ranking task. Scene length is the sum of all button times over all buttons pressed during a scene analysis, excluding any scoring buttons. Here, a button time is the time interval between the time a particular button was pressed to access the corresponding attribute's information, and the time the immediately following button was pressed, summed over all button presses for that particular button during the scene analysis.
- **Use of attribute presentation time:** The presentation time for each particular information attribute was set to the Base Presentation Time (BPT), for all three output modes, giving a common basis of comparison among the three output mode arrangements. Here, BPT was the amount of time taken for the equivalent voice message to be played back. However, a subject could spend less or more than the BPT on that attribute. The use of attribute presentation time was calculated as the percentage of BPT for that button that the subject spent before pressing the next button, averaged over all presses of that particular button during a scene. As it was possible for the subject to interrupt the output and move on to another button before the end of the BPT, this measure was used to evaluate how well a particular interface mode supported a particular subject. A value of less than 100% indicated that the subject spent less time than allowed to capture the presented information before moving to request other information, while a value of greater than 100% indicated that the subject was taking more time to think, plan, recall, and/or assimilate the information that had been presented.

Experimental design

The experiment was a partial repeated within-subject unbalanced factorial design. The two factors were: output mode (voice, text, or both (voice and text)), and cognitive style (heuristic, neutral, or analytic), blocked on order (whether a scene

was analyzed first or second in order by the subject, to account for differences due to task or interface learning effects). The design was unbalanced because there were unequal numbers of subjects in the three cognitive style classifications, and it was a partial repeated design because each subject carried out a task with only two of the three possible interface types. Equal numbers of subjects (16) were assigned randomly to each of three groups. Subjects in each group compared one of the three possible pairs of output modes (voice compared with text, voice with both, or text with both) by carrying out apartment ranking tasks in two scenes, using a different output mode for each scene.

In total there were four scenes to be analyzed. In each of the groups of 16 subjects, two scenes were assigned at random to each subgroup of eight subjects. Four of these subjects analyzed the scenes in reverse order to the other four subjects, in order to identify and correct for learning effects between the first and second scenes they analyzed. In designing the information base that was available for querying by the subjects, attribute values were varied randomly among the various apartments within realistic limits, to keep task complexity relatively constant among the subjects and tasks. Each subject's task was to evaluate his or her apartment preferences in a scene on a numerical scale from 1–10 for the four apartments in a scene; this was repeated for each of the two scenes analyzed. Although assignment of the two scenes used and their ordering for each subject was randomized in order to achieve a counterbalanced design, the order of a scene (whether analyzed first or second in order by a subject) was recorded for later use as an order blocking variable in data analysis. This allowed variability due to general task learning by the subjects to be removed from the analysis.

Subjects

As it was anticipated that subjects with task domain experience in searching for and selecting suitable living accommodation would interact differently with the interface than those with no experience, both experienced and inexperienced subjects were used to gauge the impact of domain experience on the results. However, there were not enough inexperienced subjects to carry out a complete factorial design of all mode comparisons, so data from these subjects were used in a subsidiary manner to do a limited comparison on the experience dimension.

Experienced subjects 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. Twenty-one 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 12 Analytics, five Heuristics, and 31 Neutrals, according to the Taggart and Robey (1981) classification structure. All were familiar with the use of a mouse and with the interactive use of computers; 92% had used computers extensively in their course work or in business applications, but only 55% had experience with computer-based voice output systems.

The subsidiary experiment included 22 inexperienced subjects, 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

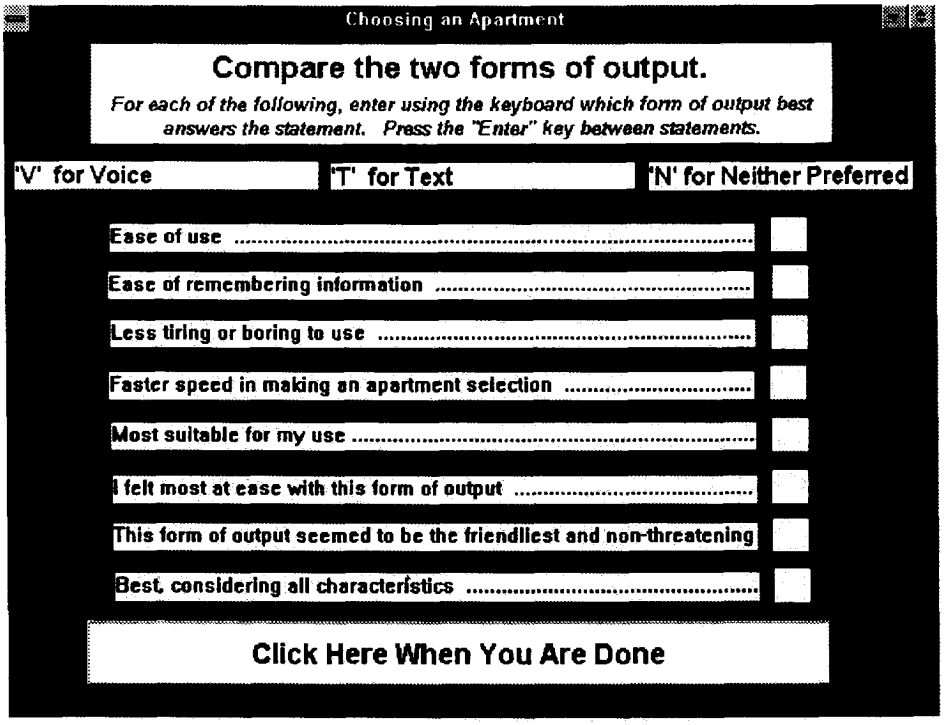


Figure 2. Interface comparison questionnaire (in this example, text is compared to voice)

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. These subjects were slightly less familiar with computer use than the MBA students; 65% had used computers extensively for course work or in business applications, 90% had previously used a mouse, and 20% had experience with computer-based voice output systems. As we did not have enough inexperienced subjects to perform a study which was complementary to the group of 48 experienced subjects, data gathered from inexperienced subjects was used only in a partial examination of Proposition 4.

Before the task, all subjects were given a short automated 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 the two apartment ranking tasks. The total time required by a subject to train and to complete the entire apartment ranking process varied from about 20 minutes to 45 minutes.

Data analysis

The data analysis reported in this section refers only to the results from the

Table 1. Interface comparison preference results for experienced subjects

	Text vs. voice			*Both vs. Text*			*Both vs. Voice*		
Question	Pref.	p	sig.	Pref.	p	sig.	Pref.	p	sig.
Ease of use	8 : 5	0.29	ns	6 : 6	0.61	ns	14 : 1	0.000	***
Remembering	9 : 7	0.40	ns	9 : 5	0.91	ns	16 : 0	0.000	***
Less tiring/boring	6 : 8	0.79	ns	8 : 4	0.93	ns	9 : 0	0.000	***
Faster	9 : 7	0.40	ns	2 : 11	0.011	**	13 : 1	0.001	***
Most suitable	9 : 5	0.21	ns	6 : 9	0.30	ns	13 : 1	0.001	***
Most at ease	7 : 5	0.39	ns	5 : 10	0.15	ns	15 : 0	0.000	***
Friendliest	4 : 4	0.64	ns	6 : 6	0.61	ns	14 : 0	0.000	***
Best overall	8 : 8	0.60	ns	8 : 7	0.70	ns	15 : 0	0.000	***

Data collected through preference questionnaire (see Figure 2)
Pref. (preference score) is in the form $m:n$, where m = number of subjects preferring first interface and n = number preferring second interface, with 16 responses. Where $m + n \neq 16$, the remainder were indifferent between the two interfaces
Sign test used for statistical comparisons
p = sig level, sig. = ns (not sig.), **(0.05 level), ***(0.001 level)
In the auxiliary experiment for inexperienced users (data not shown here), differences were not significant for any question on voice *versus* text or text *versus* both. The voice *versus* both experiment was not performed by the inexperienced users

experienced subjects. Unless indicated otherwise, two-way analysis of variance on mode and style factors, blocked on order, was used to analyze each data set. As reported in the following, F is the partial F ratio for the measure in question, and n_1 and n_2 are the degrees of freedom in the factor and the error term respectively. Recommended techniques for analyzing unbalanced designs (Appelbaum and Cramer, 1974), using a regression approach to ANOVA (Neter *et al.*, 1985) were followed throughout. Unless stated otherwise, the significance level assumed was $\alpha = 0.05$.

Output mode preferences and their 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, using a sign test pairwise comparison, is shown in Table 1. There is no significant difference in preference between the Voice or Text output modes. However, in the voice *versus* both comparisons, both was 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).

The user's final decision

To analyze the impact of output model on user decisions, a two-way analysis of variance was performed on the scores assigned to the four different apartments in each scene, with mode and apartment as factors. The results were significant for the apartment factor ($F = 18.9, 5.0, 14.7, 21.6$ respectively, with $n_1 = 3$, $n_2 = 84$, $p < 0.01$) for each of the four different scenes. On the other hand, pairwise factor

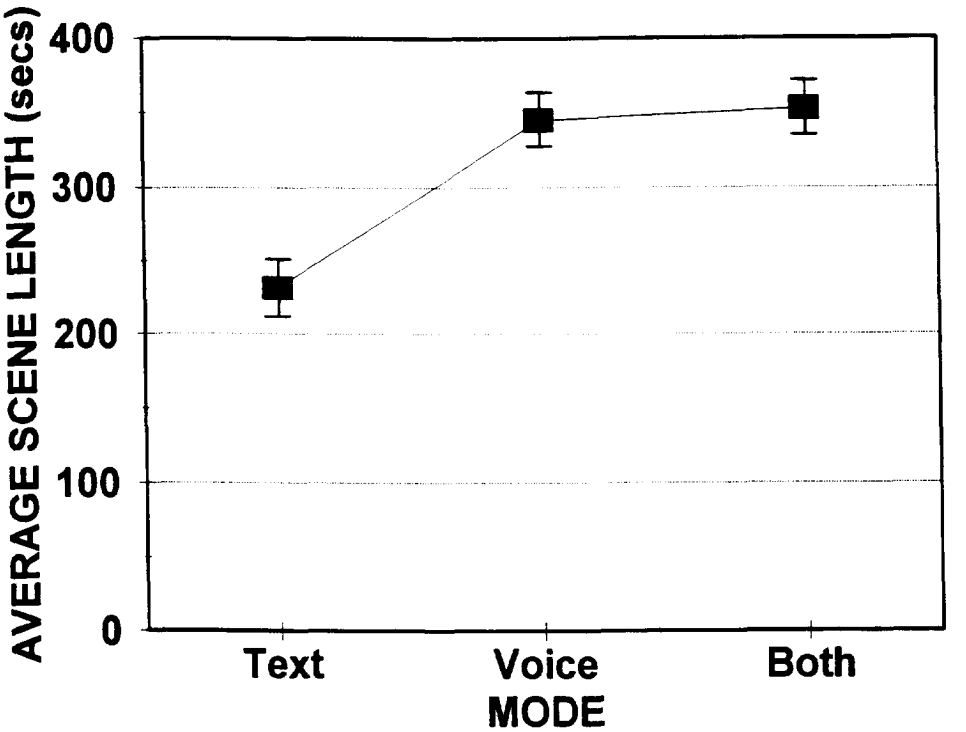


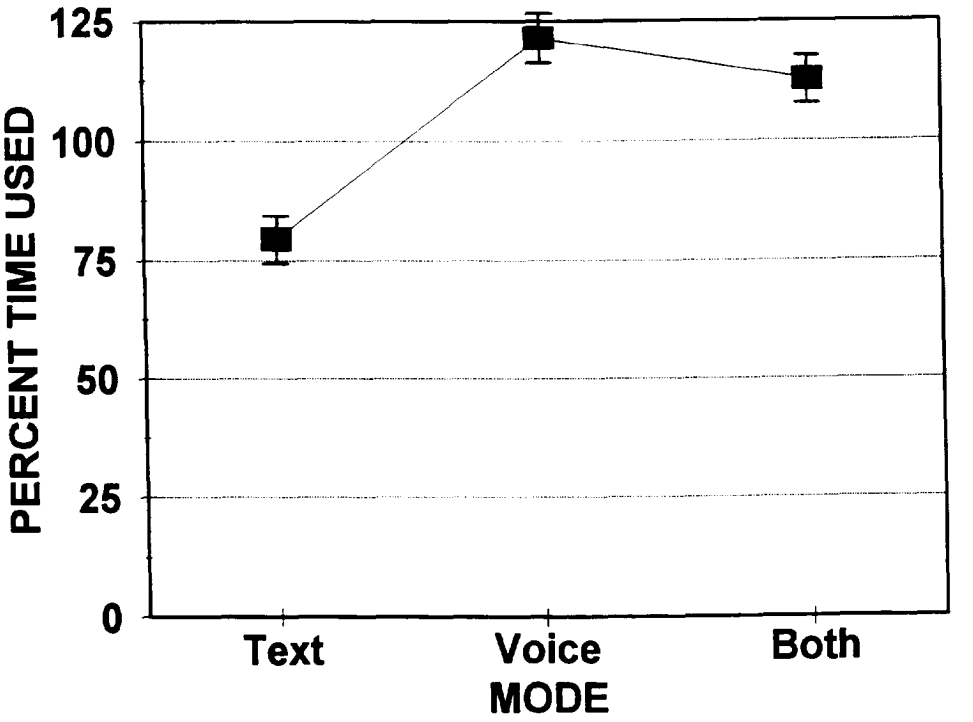
Figure 3. Output mode and average scene length

interactions and the main effect for mode were not even marginally significant in any of the scenes. This indicates that the output mode did not have a significant impact on user ratings for the apartments, and any differences observed in efficiency or user preferences during the decision-making process did not appear to have affected decision outcomes.

Output mode efficiency

Efficiency refers to the amount of time required to complete a task. Scene length, (scene completion time, excluding scoring button times) averaged over the 96 scenes evaluated by the subjects was 307 seconds, with a standard deviation of 123 seconds. Pairwise interactions and the main effect for style, on this variable were not significant. However, the main effect for mode was significant ($F = 4.84$, $n_1 = 2, n_2 = 90, p = < 0.05$), and its levels are plotted in Figure 3. Using the Tukey-Cramer method (Neter *et al.*, 1985), paired comparisons indicated that the differences between text and both and between text and voice were significant. The text mode was clearly more efficient than the other two modes evaluated.

We also analyzed the relative time taken per attribute selected, as a percentage of the time the message was presented (base presentation time for that message), to help highlight the time differences between output modes. An analysis of variance was performed on the presentation time utilization data, normalized to the number of attributes (1, 4 and 14 respectively) for the three abstraction levels.



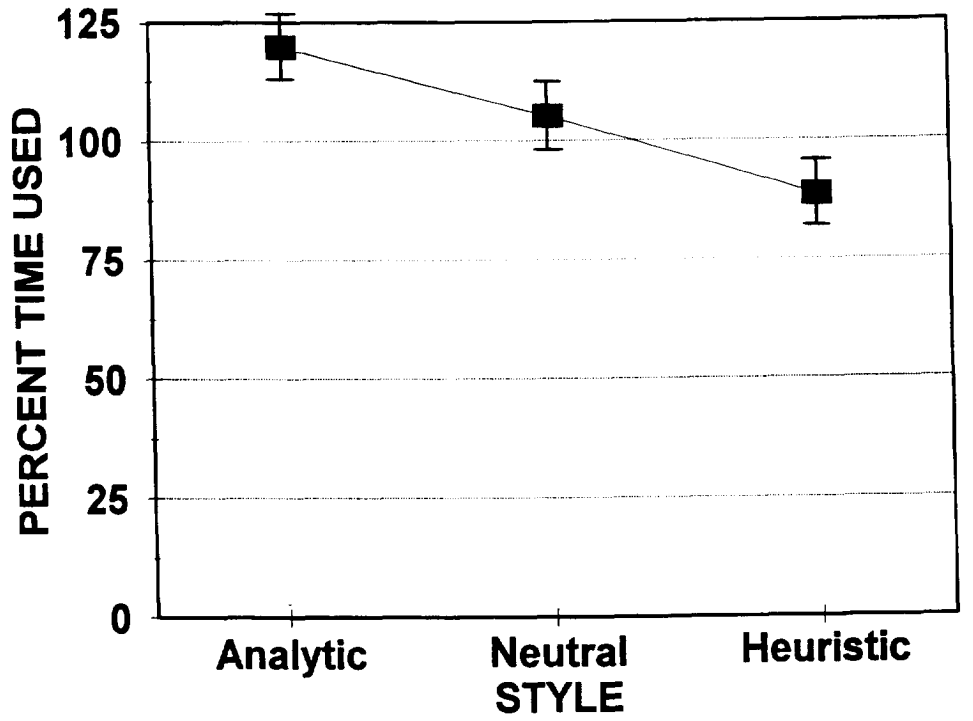


Figure 4(a). Output mode and relative presentation time per attribute reference.
Figure 4(b). Cognitive style and relative presentation time per attribute reference

Table 2. Attribute importance ranking

	Rank by importance rating		Mean number of references		Rank by mean number of refs	
Column no.	1	2	3	4	5	6
Attribute	Experienced	Inexperienced	Experienced	Inexperienced	Experienced	Inexperienced
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

Attribute abstraction levels: ** Overall (highest), *General (intermediate), All others at specific (lowest) level

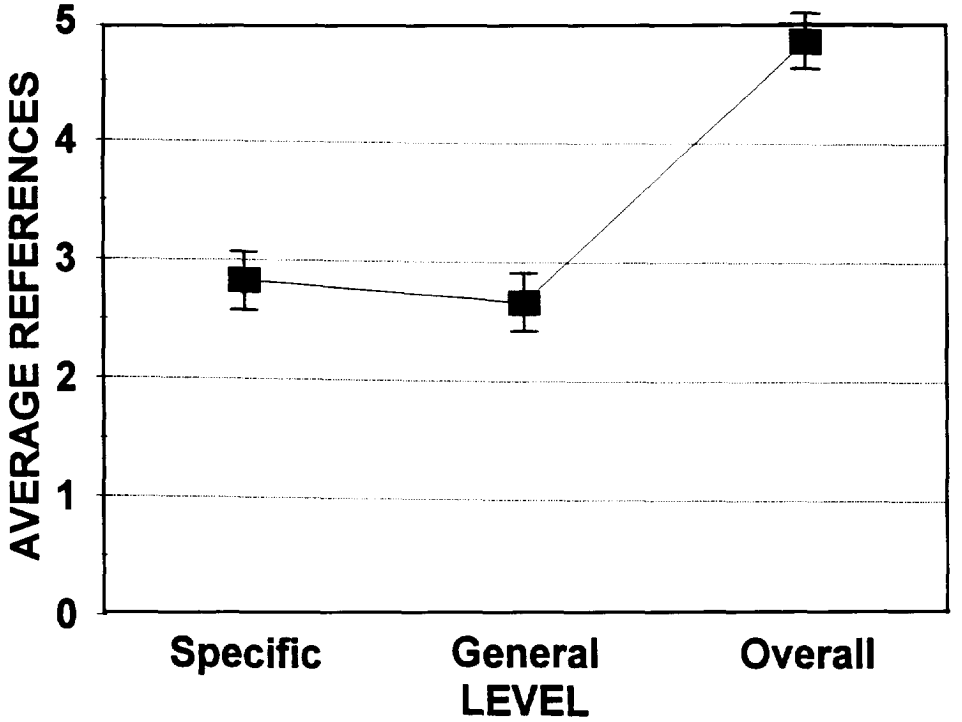


Figure 5. Abstraction level and average references per attribute

The interactions between the factors were not significant, but the main effects for mode ($F = 18.8$, $n_1 = 2$, $n_2 = 78$, $p < 0.001$), and style ($F = 5.02$, $n_1 = 2$, $n_2 = 78$, $p < 0.01$) were significant. The results of this analysis are shown in Figures 4(a) and 4(b), for mode and style main effects, respectively. Differences among the factor levels for mode in Figure 4(a) were significant, except for the differences between both and voice. All the factor level differences for style in Figure 4(b) were significant.

Usage and preference for attributes

Usage of attributes at different abstraction levels

Table 2 shows the mean frequency of access by subjects for the information attributes used in the experiment (column 3 for experienced subjects), and the resulting ranking by actual number of references (column 5 for experienced subjects). For experienced users, of the 19 possible information attributes which could be queried for each apartment, only two (overall and rate) were queried on average more than once per apartment in a scene (average greater than 4.0 for 4 apartments).

An analysis of variance on attribute selection frequency data for experienced subjects, separately for the three results obtained by averaging according to abstraction level for each subject and scene (the average number of references per attribute per scene at each level), revealed no significant interactions between mode or style; neither were their main effects significant. However, a comparison

(using one way ANOVA, blocked on order) of this measure at the three abstraction levels (overall, general, and specific) of the aggregate data was significant ($F = 56.7$, $n_1 = 2$, $n_2 = 262$, $p < 0.001$). Average references per attribute at the three abstraction levels are plotted in Figure 5. Paired differences in this plot were significant, except for the difference between general and specific. It is apparent that, although neither mode nor style explained differences in the use of abstraction level, aggregate attribute use at the highest abstraction level was significantly higher than attribute use at the two lower levels.

The average fraction of available attributes which were *not* referenced at particular levels during scene analysis were 0.12, 0.27, and 0.23 respectively, for the overall, general, and specific levels. This means that, for example, in the 96 scene analyses that were carried out by these experienced 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 12 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 Payne's paper-based manual system.

The importance of attributes at different abstraction levels

Table 2 also shows rankings of importance perceived by subjects for the information attributes used in the experiment (column 1 for experienced subjects). From the table, the rate attribute was clearly regarded as the most important by both experienced and inexperienced subjects, and such attributes as closet, shopping, and parking were ranked at the low end of the scale. However, the overall attribute was *perceived* (column 1 data) as having considerably less importance than *actual* retrievals of this attribute indicated (column 5 data).

The impact of individual differences

Preference and performance differences between different cognitive styles

As we have already mentioned, cognitive style was not a significant factor in preference for output mode or for information attributes. However, regarding performance, it was significant in measuring the use of presentation time, as shown in Figure 4(b). Figure 4(b) shows that the relative time taken per attribute selected (i.e. increased percentage of time used) by Analytics is longer than Neutrals, which in turn is longer than Heuristics. Heuristics moved on to the next attribute on average when less than 90% of the presentation time was completed, while Analytics stayed on for an average of 120% of the available presentation time.

Preference and performance differences between experienced and inexperienced subjects

The subsidiary experiment with inexperienced subjects did not include enough subjects to provide an additional factor of inexperience *versus* experience in the design. Preferences for output mode (text *versus* voice, and text *versus* both only, the only comparisons done) among inexperienced subjects was the same as for experienced subjects. Thus, domain experience appeared to have no impact on

output mode preference. Regarding preference among attributes at different abstraction levels, an analysis of variance comparison of the aggregate abstraction level data showed that experience was also not significant for this variable. To compare attribute usage differences between inexperienced and experienced subjects, the Spearman correlation coefficient was calculated between the rankings of the mean number of references for inexperienced subjects and experienced subjects (columns 5 and 6 of Table 2). It revealed an r_s of 0.78 ($p < 0.01$), a significant degree of similarity between rankings by experienced and inexperienced subjects.

Findings

In the following, we review our findings concerning the propositions stated earlier.

Proposition 1: When representing the same information, a voice and text combination will be preferred by users over either voice only or text only, and text will be preferred over voice only.

This proposition was not fully validated by the experiment. As predicted, users did have a preference for both (text and voice) over voice. But there was no significant preference between text and voice, and there was no significant preference between both (text and voice) and text. The relative preferences imply that adding voice to text did not significantly alter the perceived utility of text, but adding text to voice significantly improved its utility. A number of subjects commented that the combination of both modes helped them to remember the information better. This is in agreement with other published results (Nugent, 1982; Baggett and Ehrenfeucht, 1983).

There is an apparent non-transitivity among these results. With text ~ voice and both ~ text, one would expect both ~ voice rather than both > voice. However, when groups of people are involved in ranking, transitivity can be violated for aggregated group preferences. In our experiment there were 48 subjects, divided into three groups of 16. Each group compared two of the three interfaces. Their overall preferences are summarized in the last question of Table 1. Suppose that preferences for all three interfaces were measured for one of these groups, and eight had a preference ordering of both > voice > text, while the other eight had a preference order text > both > voice. With preference aggregation, all the subjects would favour both over voice, but there would be a half and half split between both and text and between voice and text, similar to the situation represented in Table 1.

During the experiment, when voice alone was compared to text alone, exactly the same information was presented to the subjects. The messages were short, with no opportunity for confusion, and the female voice used was clear. The only real difference would be due to the slowness of the voice interface, but this would not be directly apparent to users, who performed the apartment selection tasks separately with the two interfaces. Hence the two interfaces would appear to be roughly equivalent to users except for individual tastes or preferences (note that our results in Table 1 indicate that this is the case for the voice and text comparison

over all measures, including the speed). On the other hand, when voice is added to text in the both interface, the additional voice output could have both a positive effect (assists remembering) and negative effect (slows the reading rate) on the text output. But adding text to voice in the both interface brings only positive effects to voice output (assists remembering and increasing communication rate). This advantage is very evident to the subject who is comparing both to voice, and hence the significant preference of both over voice. This finding has important implications. For example, we can expect that adding a small text display to a telephone interface could improve the usability of a telephone answering system, but adding voice to a computer interface which already displayed text would not improve the interface significantly. Even for some video games, it may be better to replace text with voice than to add voice to text.

Proposition 2: Information access will be faster with an interface that uses text output, compared to one with voice output or with a combination of voice and text output.
As predicted, efficiency of the text interface was significantly better than with either of the other two interfaces (average scene length was less), as shown in Figure 3. There was no significant difference between both and voice interfaces, and the presentation time utilization for voice and both was at a higher percentage than text, as shown in Figure 4(a). The voice and text combination has the effect of lengthening the average time taken before the user goes on to access another attribute (as compared to text alone). This may be because the subjects' reading of the information was being disrupted (and therefore slowed) by speech, or it could indicate that some subjects were ignoring the text output and listening to voice output instead. The both interface does have an important universality characteristic, which is that it is usable by people with either hearing or visual disabilities. As it compares well with a text interface in all the aspects we measured except efficiency, this makes it a promising approach for such users.

Proposition 3: Decisions will not be affected by the voice or text output mode used in the interface when these modes contain the same information.
Our findings are in accord with this proposition. We believe this is due to the fact that exactly the same information was available through each of the three interfaces. This result cannot be generalized to other interfaces which involve graphics, images, etc., as it is difficult to provide exactly the same information through such interfaces. The interface designs we used allowed us to measure performance dependency on output mode because information content did not differ among the modes. In many cases involving multiple output media, the information presented through different media is not the same. If this is the case, decisions made may depend on the mode. For example, although images and video are often used to present information in a pleasing manner, these modes frequently need to be augmented by text or voice to ensure that users receive all the information necessary to make a rational decision. To carry out fair comparisons between such interfaces is more difficult.

Proposition 4: Given the freedom to access information directly at different abstraction

levels, experienced users will access higher level information abstractions more frequently than will inexperienced users.

Based on a comparison between results from the main and subsidiary experiments, this proposition was not upheld, because there was no apparent difference between experienced and inexperienced subjects in their indicated preferences for output mode or for information abstraction level usage. It seems that the 'inexperienced' subjects were inexperienced in apartment hunting, but not actually inexperienced in what they thought were important attributes in choosing living quarters; this made their performance more similar to that of the experienced subjects. The implication of this finding is that an interface design should be based on a common sense approach that serves both experienced and inexperienced users well.

An interesting finding concerning relative attribute importance among the various abstraction levels is in Table 2, which shows a very important distinction between 'perceived importance' and 'actual usage'. A user may perceive the importance of an attribute to his or her decision task in a different way than the actual usage, which represents his or her information access process. Although the rate attribute was both perceived as the most important and accessed most frequently, the overall attribute appeared to have less value in determining the final rating of the apartments. But it does serve the purpose of grasping the big picture as an information search starting point and reference point, and was consequently accessed more frequently on average than attributes at other abstraction levels (see Figure 5). This high level 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) received relatively low usage rankings, as well as low importance rankings. Perhaps this level of abstraction was unnecessary as there was not very much to summarize at this level. 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 an hierarchy, which is often based on abstraction levels. There are two solutions to this problem: one is to re-organize the hierarchy with different criteria such as importance rather than abstraction level, and the second is to make it convenient for the user to access any information attribute easily and directly at any level, as was possible with our interface.

Proposition 5: Analytic individuals will take more time to make decisions than will heuristic individuals.

Our results (see Figure 4(b)) are in accord with this proposition, for relative presentation time per attribute reference. However, an interesting point is that, in terms of scene length, cognitive style was *not* significant. It appears that, although Heuristics spent less time on each attribute, they looked at more attributes before reaching the final decision. In contrast, Analytics spent more time on each attribute, made a more careful evaluation and looked at fewer but more relevant (to them) attributes in order to reach the final decision. Thus, the overall time spent for information gathering for different cognitive styles was roughly the same. The lesson for multimedia interface designers is that the interface

should not, inadvertently or otherwise, favour one style over another because cognitive style may affect how the interface will be used. For example, putting pressure on the user to assimilate information as quickly as possible may degrade user performance if the user's personality is not suited to that style of information gathering. The implication is that output mode and data display speed should be under the control of the user as much as possible.

Discussion

Direct comparisons between voice and text output modes are not simple, because too many variables may affect the results. In our study we tried to make pure comparisons, using the same information presented in different modes. However, we cannot isolate the context from the psychological impact. For instance, a warm welcome voice is more pleasant than a text notice. But a robot-like voice may not be any better than text alone and may, in fact, be worse from the user's perspective. Text information may be better than voice information if there is the chance for possible errors in understanding by the user, especially since text can be re-scanned easily, while the sequential nature of voice makes it more difficult to play back. On the other hand, combining both text and voice can aid in learning and remembering information. Further research is needed to expand these comparisons to a much more broad setting, with a variety of decision variables.

The application of abstraction levels in interfaces is difficult to investigate, and may introduce constraints which interfere with user preferences and choices. In our study, we tried to give users freedom to access information at any abstraction level without extra effort. However, the abstraction levels we introduced did not necessarily coincide with users' perceptions of importance levels, as we discovered in our experiments. Further research is needed to investigate usage and access patterns among different levels of abstraction and importance.

Acknowledgements

We wish to thank the reviewers, who made a number of suggestions that improved the quality of this paper. This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

References

Anderson, J.R. (1985) *Cognitive Psychology And Its Implications* WH Freeman

Appelbaum, M.I. and Cramer, E.M. (1974) 'Some problems in the nonorthogonal analysis of variance' *Psychol. Bull.* **81**, 335–343

Archer, N.P. and Kao, D. (1993) 'An empirical study of abstraction in conceptual model design' *ASAC Banff Conf. (Information Systems) Proc.*

Baggett, P. and Ehrenfeucht, A. (1986) 'Encoding and retaining information in the visuals and verbals in an educational movie' *Educational Comm. Tech. J.* **31**, 3, 23–32

Bariff, M.L. and Lusk, E.J. (1977) 'Cognitive and personality tests for the design of management information systems' *Management Sci.* **23**, 820–829

Batra, D. and Davis, J.G. (1989) 'A study of conceptual data modelling in database design: similarities and differences between expert and novice designers', in **DeGross, J.L.,**

Henderson, J.C. and Konsynski, B.R. (eds) *Proc. 10th Int. Conf. Information Systems* ACM Press, 91–99

Benbasat, I. and Taylor, R.N. (1978) 'The impact of cognitive style on information system design' *MIS Quarterly* **2**, 43–54

Blaylock, B.K. and Rees, L.P. (1984) 'Cognitive style and the usefulness of information' *Decision Sci.* **15**, 75–91

Bly, S.A., Harrison, S.R. and Irwin, S. (1993) 'Media spaces: bringing people together in a video, audio, and computing environment' *Comm. ACM* **36**, 28–47

Bond, A.H. and Soetarman, B. (1988) 'Multiple abstraction in knowledge-based simulation', in **Henson, T. (ed)** *Artificial Intelligence And Simulation* Society for Computer Simulation, 61–66

Chalfonte, B.L., Fish, R.S. and Kraut, R.E. (1991) 'Expressive richness: comparison of speech and text as media for revision' *CHI '91 Conf. Proc.* Addison Wesley, 21–26

Couger, J.D., Higgins, L.F. and McIntyre, S.C. (1993) '(Un)structured creativity in information systems organizations' *MIS Quarterly* **17**, 441–461

Davis, D.L., Barnes, J.H. Jr. and Jackson, W.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, M.J. and Wallace, W.A. (1992) 'The effects on decision task performance of computer synthetic voice output' *Int. J. Man-Machine Studies* **36**, 65–80

Guindon, R., Krasner, H. and Burtis, B. (1987) 'Cognitive processes in software design: activities in early, upstream design' in **Bullinger, H.J. and Shackel, B. (eds)** *Proc. Interact '87*, Elsevier, 383–388

Hodges, M.E. and Sasnett, R.M. (1993) *Multimedia Computing: Case Studies From Project Athena* Addison Wesley

Huber, G.P. (1983) 'Cognitive style as a basis for MIS and DSS designs: much ado about nothing? *Management Sci.* **29**, 567–579

Miller, G.A. (1956) 'The magic number seven plus or minus two: some limitations on our capacity to process information' *Psychol. Review* **63**, 81–97

Montazemi, A.R. and Wang, S. (1989) 'The effects of modes of information presentation on decision-making: a review and meta-analysis' *J. Management Info. Systems* **5**, 3, 101–127

Myers, I., Briggs and McCaulley, M.H. (1985) *Manual: A Guide To The Development And Use Of The Myers-Briggs Type Indicator* Consulting Psychologists Inc.

Neter, J., Wasserman, W. and Kutner, M.H. (1985) *Applied Linear Statistical Models* (2nd ed.) Irwin

Nielsen, J. (1993) *Usability Engineering*, Academic Press

Norman, K.L., Weldon, L.J. and Shneiderman, B. (1986) 'Cognitive layouts of windows and multiple screens for user interfaces' *Int. J. of Man-Machine Studies* **25**, 229–248

Nugent, G. (1982) 'Pictures audio and print: symbolic representation and effect on learning' *Educational Comm. Tech. J.* **30**, 3, 163–174

O'Keefe, R.M. and Pitt, I.L. (1991) 'Interaction with a visual interactive simulation, and the effect of cognitive style' *European J. Operational Research* **54**, 339–348

Ossher, H.L. (1987) 'A mechanism for specifying the structure of large, layered systems', in **Shriver, B. and Wegner, P. (eds)** *Research Directions in Object-Oriented Programming* MIT Press, 219–252

Payne, J.W. (1976) 'Task complexity and contingent processing in decision making: a replication and protocol analysis', *Org. Behavior & Human Performance* **16**, 366–387

Ramaprasad, A. and Mitroff, I.I. (1984) 'On formulating strategic problems' *Academy of Management Review* **9**, 597–605

Sipior, J.C. and Garrity, E.J. (1992) 'Merging expert systems with multimedia technology' *Data Base* (Winter); 45–49

Staggers, N. and Norcio, A.F. (1993) 'Mental models: concepts for human–computer interaction research' *Int. J. of Man-Machine Studies* **38**, 587–605

Streeter, L.A. (1988) 'Applying speech synthesis to user interfaces' in **Helander, M. (ed)** *Handbook Of Human-Computer Interaction* Elsevier, 321–343

Taggart, W. and Robey, D. (1981) 'Minds and managers: on the dual nature of human information processing and management' *Academy of Management Review* **6**, 187–195

Todd, P. and Benbasat, I. (1993) 'An experimental investigation of the relationship between decision makers, decision aids and decision making effort' *INFOR* **31**, 80–100

Umanath, N.S., Scamell, R.W. and Das, S.R. (1990) 'An examination of two screen/report design variables in an information recall context' *Decision Sci.* **21**, 216–240

Vessey, I. (1991) 'Cognitive fit: a theory-based analysis of the graphs versus tables literature' *Decision Sciences* **22**, 219–240

Zmud, R.W. (1978) 'Individual differences and MIS success: a review of the empirical literature' *Management Sci* **25**, 966–979

Appendix 1: example of data used in study

Table A1: data for one apartment

Attribute*	Data
Overall	One bedroom basement apartment, rather low price, 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	20-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.

*Amount of indentation is related to abstraction level, with 'Overall' at the highest level

Received June 1996; accepted September 1996