aval Research Laboratory
ıshington, DC 20375-5000

NRL Report 9372

Dual-Task Performance as a Function of Presentation Mode and Individual Differences in Verbal and Spatial Ability

LISA B. ACHILLE, ASTRID SCHMIDT-NIELSEN, AND LINDA E. SIBERT

Human Computer Interaction Laboratory
Information Technology Division

January 31, 1992

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 31, 1992	Final

4. TITLE AND SUBTITLE

Dual-Task Performance as a Function of Presentation Mode and Individual Differences in Verbal and Spatial Ability

5. FUNDING NUMBERS

PE - 601153N
RR015-09-41

6. AUTHOR(S)

Lisa B. Achille, Astrid Schmidt-Nielsen, and Linda E. Sibert

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL Report 9372

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The effectiveness of alternative display formats as a function of individual differences in verbal and spatial abilities was evaluated in a dual-task paradigm. Tasks consisted of two-dimensional tracking and a classification task in which items were presented as text, speech, or icons. Spatial ability was correlated with performance on the tracking task both for single task and for dual task in combination with the various presentation modes of the classification task. Verbal ability was not consistently correlated with performance on any of the tasks. Significant individual differences in dual-task performance were found, and individuals were highly consistent with themselves across different presentation modes. Classification task performance is compared for the three presentation modes singly and in combination with the tracking task. Dual-task classification was slower than single-task classification for the visual modes, but there was no increase in reaction time between single- and dual-task performance for speech. In the dual-task conditions, the largest tracking performance decrements were found for the text condition, with smaller decrements for speech, and smallest decrements for iconic presentations. Issues related to time-sharing ability and strategies are also discussed.

14. SUBJECT TERMS

Display format
Dual-task performance
Individual differences
Time-sharing ability
Presentation mode

15. NUMBER OF PAGES 25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

CONTENTS

DUAL-TASK PERFORMANCE AS A FUNCTION OF PRESENTATION MODE AND INDIVIDUAL DIFFERENCES IN VERBAL AND SPATIAL ABILITY

INTRODUCTION

The experiments described in this report were designed to evaluate the effectiveness of alternative display formats as a function of individual differences in verbal and spatial abilities. This is part of an ongoing effort to investigate cognitive factors that influence the use of complex, high workload systems. Complexity in the context of the work reported here refers to a tendency toward requiring system users to perform multiple subtasks within a single system. The goal of the larger effort is to facilitate the use of demanding interfaces through performance metrics and design intervention.

The results of display research are usually reported on the basis of average performance for a group of subjects and do not focus on individual performance differences. An examination of factors influencing human effectiveness in dealing with displayed information, especially in high workload situations where individual differences in skill levels and strengths can most affect overall performance, can help determine the potential benefits of adapting displays to individual skills or cognitive strength areas. Factors influencing human effectiveness include: the relationship between cognitive skills and information presentation modes, the relationship between component tasks in multitask situations, compatibility between the presentation mode and the nature of the information presented, and variation in cognitive workload as a function of individual differences in cognitive skills. These effects could be realized either as a difference in overall performance or a difference in dual-task performance, either in strategies or in time sharing between the two tasks. In addition to evaluating the potential value of adapting systems to individual needs, the results of individual difference analyses might also contribute to the generation of personnel selection guidelines or specialized training programs.

Several studies have shown that large individual differences exist in general verbal and spatial abilities (e.g., Refs. 1 through 5) and that there are different types of spatial ability. There is also evidence that cognitive abilities have been predictive of some computer interactive performance. Egan and Gomez [6] showed that spatial ability is positively correlated with the learning of text editing skills in terms of time spent and errors made. Peters, Yastrop and Boehm-Davis [7] linked individual differences in perceptual speed and spatial scanning abilities [8] with information retrieval performance when database format is varied, suggesting the existence of other relationships in which human information processing might interact with task characteristics. Given that there are considerable differences in cognitive skill levels across individuals and that the number of options for display design and complexity continues to increase, any relationships that are found between individual skills and performance may help determine display effectiveness for a particular task.

In addition to individual differences in verbal and spatial skills, the ability to integrate or coordinate activities in multitasking situations is an important factor in overall performance. Several investigators have looked at individual differences in the time sharing ability under dual-task conditions (e.g., Refs. 9 through 11). Damos, Smist, and Bittner [10] found that different subject strategies influenced the effectiveness of dual-task performance. Forrester [11] found individual differences in subjects' ability to cope with differences in difficulty level. Ackerman, Schneider, and Wickens [12] examined a variety of methodological issues in the way data from

Manuscript approved October 1, 1991.

dual-task experiments are analyzed and interpreted, and they concluded that while the existence of a time sharing ability could not be rejected, methodological issues in previous studies also precluded strong support for such an ability. After taking single-task performance into account, Yee, Hunt, and Pellegrino [13] found individual differences in the ability to coordinate information from different sources to accomplish a task.

Many military and civilian systems require the user to attend and respond to more than one set of stimuli at the same time. For these experiments, a dual-task paradigm was chosen to represent the effect of multiple task demands in using complex systems. A tracking task and a classification/decision task were selected to limit the experimental domain to broadly defined areas of spatial and verbal processing. The tracking task was chosen to represent a class of spatial skills that are by nature analog and important to guiding vehicles, tracking targets, and other activities common to command and control tasks that involve keeping track of objects in space [14]. The classification task was chosen to represent the class of activities important in evaluating the nature of incoming information, selecting among alternative targets, and so forth.

The tracking task used a moving target with random direction changes to require continuous attention to the task and to prevent automation of the task. The classification task required the subject to hold two category names in memory for the duration of each trial and to use these to classify the objects presented at regular intervals. The classification task was used to vary the presentation mode of the items to be classified. Three presentation modes were selected as being representative of possible options available to interface designers — icons, text, and speech. The modes were selected to allow for comparisons across sensory presentation mode and mental codes for the presented items (see Table 1). Icons are visually presented and use a visual code, whereas text is also visually present but uses a verbal code, and speech is auditorily presented and uses a verbal code. Multiple resource theory [15-16] holds that competition among tasks should be less when the two tasks use different mental resources (e.g., auditory vs visual) and also that performance on a given task will be better if the stimulus, central processing, and output demands are compatible (e.g., visual-spatial coding with manual output or auditory-verbal coding with spoken output). In terms of central processing, the way in which the two tasks interact for different presentation modes may vary with individual differences. Those with high spatial ability may find the icons easier to process, whereas people with low spatial ability may have more difficulty when both inputs are visual. Some conflict can be expected from controlling two different manual outputs, but this is constant across all conditions.

Table 1 — Resource Demands for the Tasks in Experiment 1

	Classification Task			Tracking Task
	Icons	Text	Speech	Easy/Hard
INPUT	Visual	Visual	Auditory	Visual
PROCESSING CODE	Spatial	Verbal	Verbal	Spatial
OUTPUT	Manual	Manual	Manual	Manual

The main components of information processing tasks in real-world systems are still largely verbal and, in fact, text only is often used. One exception is location information, which is generally coded spatially with symbols, particularly for geographic displays. To separate the verbal from the spatial components as much as possible, the location aspect was omitted from the classification task while retaining the symbolic aspects for iconic displays. Speech was included because it is currently underutilized as a designers' option in most complex systems and because it provides a nonvisual, verbal mode.

There is some indication that high verbal ability subjects classify items faster overall than those with lower verbal abilities. The work of Goldberg, Schwartz, and Stewart [17] shows a correlation between high verbal skills and fast classification ability in a set of tasks progressing in complexity from determining physical identity to name identity to semantic class identity. They found an increased divergence in reaction time (RT) between low-verbal subjects and high-verbal subjects. As the lexical decision required became more complex, the advantage of the high-verbal subjects increased. Given the semantic nature of our classification task, we might expect high-verbal ability to be correlated with relatively high performance on classification in general, regardless of presentation mode. It is possible that the high-verbal subjects may simply be better at classifying speech and text when presented in the single-task condition, without an additional advantage when combined with the spatial task.

METHOD

General Procedure

The experiment consisted of tests of verbal and spatial abilities in one session followed by either two or three sessions of single-task and dual-task testing on the two experimental tasks: a tracking task and a classification task. The tracking task had two levels of difficulty (easy tracking and difficult tracking) and the classification task had three presentation modes: icons (the Snodgrass and Vanderwart pictures [18]), text (the printed names of the items), and speech (spoken names of the items produced by a speech synthesizer). The tracking task consisted of using a mouse to try to keep a cursor on the target, a black circle that could move in eight random directions within a rectangular area of the screen. The target changed direction every 2.5 s, pausing briefly before each change. When the target reached the edge of the rectangle, it appeared to "bounce" back. The classification task required the subject to hold two category names in memory for the duration of each 2.5 min trial. The categories changed from trial to trial to reduce learning effects, and the task was to decide whether or not a presented item was a member of one of the target categories and to push one of two buttons indicating a yes or no decision. Items were presented every 2.5 s, timed to coincide with the direction changes on the tracking task.

Subjects were tested individually, and all parts of the experiment were controlled by a Macintosh MacPlus computer. Each subject was seated at approximately 20 in. from the screen. The MacPlus was equipped with a standard mouse and mousepad, and a custom-designed electronic reaction time/response recorder with response keys labelled YES and NO. Subjects were instructed to use their preferred hand to control the mouse, which was the response device for the tracking task, and to use the nonpreferred hand on the RT recorder to respond to the classification task.

The tracking task area was presented in the upper left-hand section of the Macintosh display in a window 3.47 in. wide and 3.80 in. high. The target was a black circular area of 20 pixel diameter (0.28 in.) for the easy tracking and 10 pixel diameter (0.14 in.) for the hard tracking. The classification task items for the two visual presentation modes — icons and text —were presented in a 3.42 X 3.42 in. window in the upper right-hand section of the screen. The icon presentation mode used the Snodgrass, Smith, Feenan, and Corwin [19] electronic picture set. The individual pictures varied in size and all fit within the window. For the text presentation mode, the names of the items were presented in capital letters using 12 point Geneva font. Item names varied in length from 3 to 12 letters. Synthesized speech was used in the first experiment because it is representative of computer voice output. It also provided a controlled, repeatable set of stimuli. The speech stimuli were synthesized using an early version of a synthesis system that was under development at NRL. The lists were tape recorded at one word every 2.5 s and presented to the listeners via Realistic Pro-60 headphones using an Otari model MX5050BQII reel-to-reel tape recorder.

Subjects were instructed to divide their attention as evenly as possible between tracking and classification in the dual-task condition.

Pretests

The first session consisted of three brief tests of verbal and spatial abilities. All pretest materials were adapted using a Hypercard application to be automatically administered and scored on a Macintosh computer. A vocabulary test was selected to test verbal ability because vocabulary is known to be highly correlated with general verbal ability as well as with other measures of verbal ability [3]. Items from the Educational Testing Service (ETS) vocabulary tests V-1, V-2, V-3, and V-4 [8] were combined to generate the verbal abilities pretest. Two tests of spatial ability—a mental rotation test (which tests spatial relations skills) and a mental paper folding test (which tests spatial visualization skills)—were selected because no single test of spatial ability exhibits the high correlations with most other spatial ability measures [20] as vocabulary does with verbal ability measures. The ETS Card Rotations Test (S-1) and the ETS Paper Folding Test (VZ-2) were used with additional items generated by the experimenters to make the tests longer. All three tests were timed, and more problems were supplied than could be completed in the allotted time. The time limits were 9 min for vocabulary, 6 min for mental rotation, and 10 min for mental paper folding.

Tracking Task

Tracking was performed by using a standard mouse set at the slowest speed to control the cursor. The target started at a random location in the tracking window and could move in eight possible straight line directions. If the target reached one of the boundaries of the rectangle, it appeared to bounce back into the tracking area. The target paused briefly every 2.5 s and changed direction at random when it began to move again. The beginning of each new target movement was timed to coincide with the presentation of the item for the classification task in the dual-task condition. Time on target was used as the measure of tracking performance. Scoring was based on an invisible extended target region (30 by 30 pixels square for the easy level and 20 by 20 pixels square for the difficult level) in which the visible target was centered. The cursor was counted as being on target if the center of the cursor was within the target region. The data collection program scored tracking performance by checking the cursor location periodically (30 times for each direction change, or each time an item was presented for the classification task) and awarding a point whenever the cursor was within the target region. The maximum possible tracking score was 1800 for both easy and difficult tracking, and a minimum score of about 12 occurred by chance if the cursor was left in one position throughout the trial.

Classification Task

For the classification task, subjects were given two target category labels (e.g., toys and furniture) and asked to identify each item that was presented as either belonging or not belonging to one of the target categories. Subjects pressed the YES key on the response time recorder for target items and the NO key for nontarget items. The target category labels changed from trial to trial and were presented to the subjects on the initial screen before each trial and were also read aloud by the experimenter to ensure that the subject attended to the target categories for each trial. Thus, there was also a memory load in that the subjects were required to remember target categories for the duration of each trial. Each trial consisted of 60 items (12 targets and 48 nontargets) presented at a rate of one item every 2.5 s.

Items for each list were taken from the 150 items in the Snodgrass et al. [19] electronic picture set, with two sets of 6 target items selected from each of the two selected target categories and the 48 nontargets chosen randomly from the remaining items, except that, as far as possible, items that might easily be misclassified as targets for a given list were rejected and replaced (e.g., if one target category was toys, truck and airplane were excluded).

EXPERIMENT 1

Subjects

Two separate groups of subjects were tested. Forty-nine NRL employees volunteered to participate without compensation. Complete data were collected from 43 of these, 15 females and 28 males. These included clerical, administrative, and technical personnel. Forty-four college students from the University of Maryland undergraduate psychology department subject pool volunteered to participate for extra course credit. Complete data were collected from 40 of these, 30 females and 10 males. The total number of subjects with complete data was 83.

Design

Experiment 1 was conducted in four sessions for the NRL subjects and three for the University of Maryland subjects (session four was omitted due to time constraints on subject participation). Rest breaks were given between sessions if multiple sessions were scheduled on one day. Verbal and spatial ability tests were given in session one. Session two was considered to be a practice session. At the beginning of the second session, subjects were familiarized with the speech synthesizer by listening to two presentations of all 150 item names while following along on a printed list of the words. The remainder of session two and the following session(s) consisted of single- and dual-task testing with the test orders shown in Table 2. To control for the effects of practice and/or fatigue, the order of the presentation modes for the classification task was balanced across three groups of subjects, with subjects assigned randomly to groups. Within task conditions, easy tracking always preceded hard tracking.

Results

Pretest Scores

The scores on all three pretests were corrected for guessing based on the number of response alternatives. There was a significant, though not large, correlation between the mental rotation test and the mental paper folding test, $r = 0.439$, $p < 0.001$, suggesting that there was some relationship between the two tests but that they were also measuring different things. A combined spatial ability score was calculated by converting the scores on the two tests to z-scores, averaging them, and then converting the averaged scores back to the same scale as the rotation test. A small but significant correlation also existed between spatial ability scores and verbal ability (vocabulary score), $r = 0.345$, $p < 0.01$. Figure 1 shows the relationship between verbal (VERB) and spatial (VIS) ability scores. There was a wide range in both verbal and spatial ability scores, and high spatial ability scores could be associated with either high or low verbal ability scores, but there were no cases of low spatial ability associated with the high verbal ability scores.The mean score for verbal ability was 33.7, with a standard deviation of 23.0. The mean score for spatial ability was 56.3, with a standard deviation of 19.5. The NRL subjects and the university students had significantly different scores on both verbal ability, $t = 3.67$, $p < 0.001$ and on spatial ability, $t = 2.12$, $p < 0.05$. The mean verbal scores were 42.0 for NRL and 24.7 for the students, and the mean spatial scores were 60.6 for NRL and 51.7 for the students.

Presentation Mode Effects

Table 3 shows the average scores for the NRL subjects and the University of Maryland subjects for single- and dual-task performance on the two tasks. The average scores on the classification task were almost identical on each of the presentation mode conditions for the two groups. The pattern of results for the tracking task was also very similar for the two groups, except that the NRL subjects had somewhat higher tracking scores in all conditions. Since the NRL subjects also had higher spatial ability scores, higher tracking scores were to be expected based on the high correlation between spatial ability and tracking performance, which will be discussed in more detail in the section on individual differences. Because of the high degree of similarity of the results for the two groups on single- and dual-task performance, the data from the 43 NRL subjects and the 40 University of Maryland subjects were combined in the overall data analysis.

Table 2 — Test conditions for Experiment 1

GROUP 1	GROUP 2	GROUP 3
	SESSION 2	
	Single-Task Classification	
Icon	Text	Speech
Text	Speech	Icon
Speech	Icon	Text
	Single-Task Tracking	
Easy	Easy	Easy
Easy	Easy	Easy
Hard	Hard	Hard
	Rest Period	
	Dual-Task Classification and Tracking	
Text w/easy	Speech w/easy	Icon w/easy
Text w/hard	Speech w/hard	Icon w/hard
Speech w/easy	Icon w/easy	Text w/easy
Speech w/hard	Icon w/hard	Text w/hard
Icon w/easy	Text w/easy	Speech w/easy
Icon w/hard	Text w/hard	Speech w/hard
	SESSION 3	
	Single-Task Classification	
Speech	Icon	Text
Icon	Text	Speech
Text	Speech	Icon
	Single-Task Tracking	
Easy	Easy	Easy
Hard	Hard	Hard
	Dual-Task Classification and Tracking	
Icon w/easy	Text w/easy	Speech w/easy
Icon w/hard	Text w/hard	Speech w/hard
Text w/easy	Speech w/easy	Icon w/easy
Text w/hard	Speech w/hard	Icon w/hard
Speech w/easy	Icon w/easy	Text w/easy
Speech w/hard	Icon w/hard	Text w/hard

Table 2 (Cont'd) — Test conditions for Experiment 1

GROUP 1	GROUP 2	GROUP 3
	SESSION 4	
	Dual-Task Classification and Tracking	
Text w/easy	Speech w/easy	Icon w/easy
Text w/hard	Speech w/hard	Icon w/hard
Speech w/easy	Icon w/easy	Text w/easy
Speech w/hard	Icon w/hard	Text w/hard
Icon w/easy	Text w/easy	Speech w/easy
Icon w/hard	Text w/hard	Speech w/hard
	Single-Task Classification	
Text	Speech	Icon
Speech	Icon	Text
Icon	Text	Speech
	Single-Task Tracking	
Easy	Easy	Easy
Hard	Hard	Hard

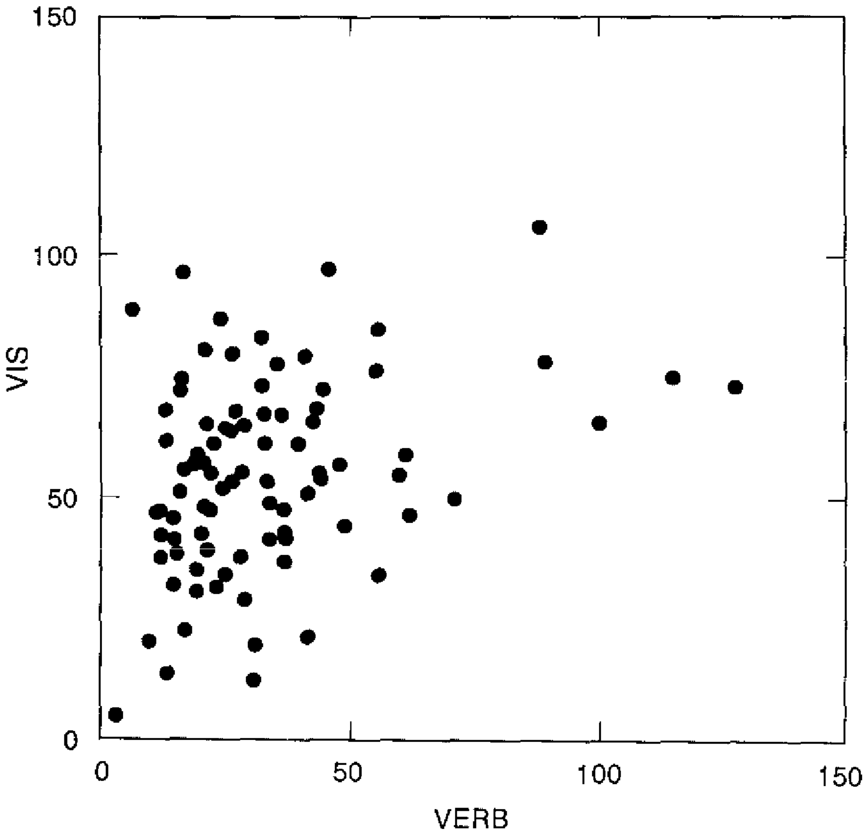

Fig. 1 — Relationship of verbal (VERB) and spatial (VIS) abilities

Table 3 — Comparison of University of Maryland (UMD) and NRL subjects' performance for tracking and for classification tasks

TRACKING SCORES

Tracking Level	Easy		Hard	
Subject Group	UMD	NRL	UMD	NRL
Single tracking	818	920	647	764
Tracking w/Icons	702	836	545	701
Tracking w/Text	617	753	447	589
Tracking w/Speech	686	806	500	619

CLASSIFICATION RT (ms)

Test Condition	Icons		Text		Speech	
Subject Group	UMD	NRL	UMD	NRL	UMD	NRL
Single Task	496	523	820	843	682	695
W/easy Tracking	661	628	1358	1352	823	824
W/hard Tracking	637	597	1418	1388	857	834

Tracking Task — Figure 1 shows single- and dual-task tracking score means. Because the scores for easy tracking and for hard tracking would be expected to be different because they were based on different sized target areas, separate analyses were performed for easy and hard tracking. Repeated measures analysis of variance showed significant effects of treatment conditions for both easy tracking, $F(3,246) = 105.0$, $p < 0.001$, and for hard tracking, $F(3,246) = 140.1$, $p < 0.001$. Multiple comparison tests were carried out using the Tukey HSD test [21] ($p < 0.01$). Single-task performance was significantly better than dual-task performance for all three presentation modes on the classification task, for both easy and hard tracking conditions. Performance on easy tracking was significantly better when combined with icons or speech than with text presentation mode on the classification task, but icons and speech did not differ significantly. Performance on hard tracking was significantly better when combined with icons than with speech, and both icons and speech were significantly better than text.

Classification Task — The classification task used two dependent measures: RT and percent errors. Figure 2 shows single- and dual-task classification performance. The RT scores for each subject were the mean RTs for the correct responses in each trial. There was a typographical error in the program for the single-task speech condition, and the RTs for single-task speech were estimated* based on other information for those trials. To compensate for nonhomogeneity of

*A typographical error in the reaction time program for single-task speech caused the timing device to be polled one second earlier than for the other conditions. This meant that longer responses were not scored for this condition. However, any responses that occurred between the time the device was polled and reset and the time that the timer was started for the next item could be tallied and counted even though the actual reaction time and the correctness of the response could not be obtained. This made it possible to calculate a slow response ratio — the ratio of long responses (those that occurred after polling) to total (long plus short) responses. A prediction formula was then derived based on the very high correlations between reaction times and the slow response ratio for the other two speech conditions. The correlations between measured reaction time and slow response ratio were 0.905 for speech with easy tracking and 0.931 for speech with hard tracking. The regression equations for predicting reaction time from slow response ratio were $RT = 686.1 + 4.51*SR$ for speech with easy tracking and $RT = 697.1 + 4.33*SR$ for speech with hard tracking. The maximum difference in reaction times predicted from these two equations is less than 7 ms, and the equation used to obtain estimated reaction time from slow response ratio for single-task speech, determined by averaging the two, was $RT = 601.6 + 4.42*SR$.

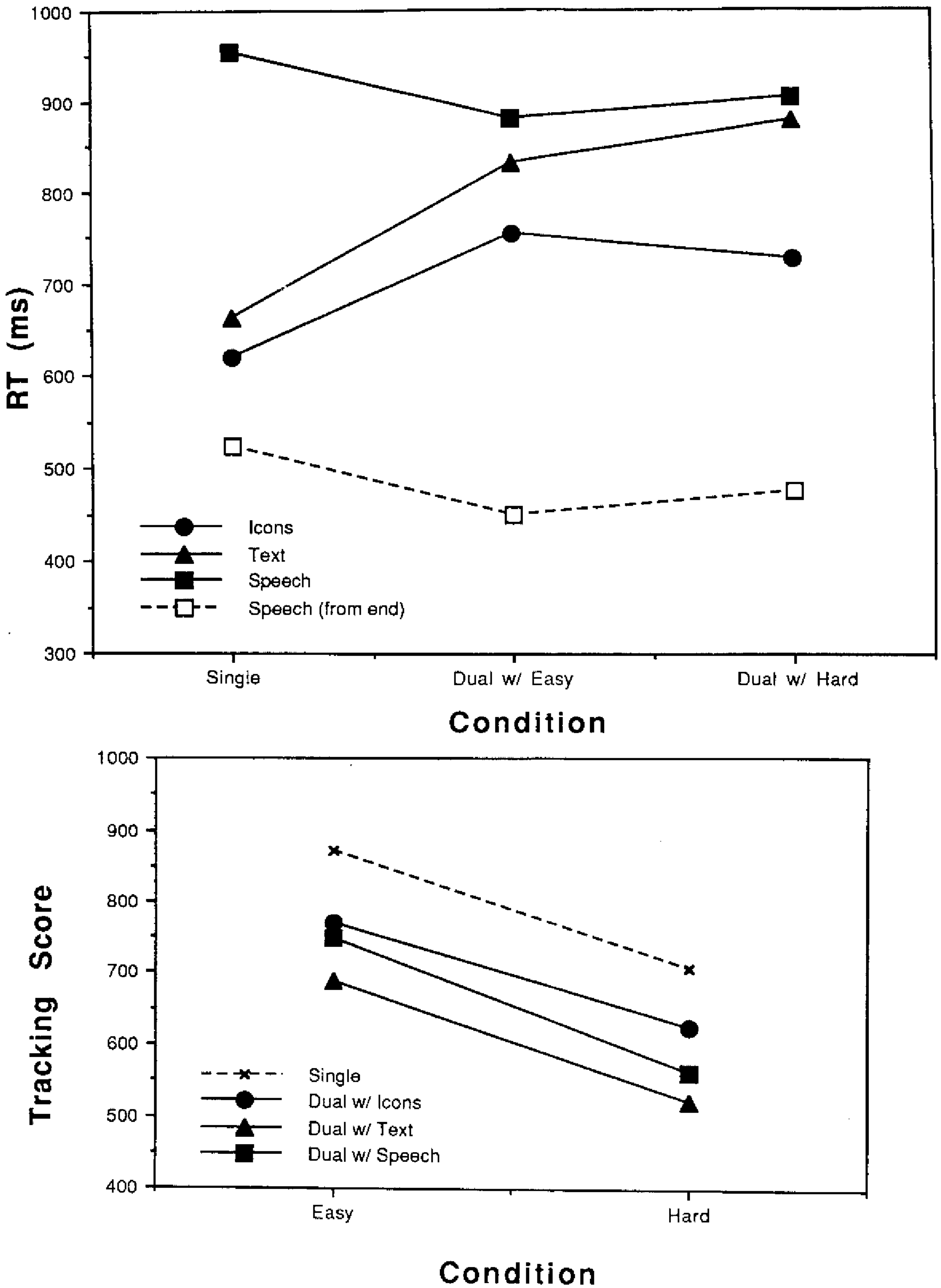

Fig. 2 — Classification and tracking performance for Experiment 1. To facilitate comparisons with the icon and text modes, classification RT for the speech presentation mode is shown twice, from word onset (solid line) and from word offset (broken line).

variance, the two-way repeated measures analysis of variance was performed using log-transformed RT scores. Presentation mode had a significant effect, $F(2,164) = 190.9, p > 0.001$, as did task condition, $F(2, 164) = 44.9, p > 0.001$, and there was a significant interaction, $F(4, 328) = 70.0, p < 0.001$.

Multiple comparison tests on the RT data were carried out using the Tukey HSD test ($p < 0.01$). For icon and text presentation modes, single-task performance was significantly better than both of the dual-task conditions, but the differences between dual task with easy tracking and dual task with hard tracking were not significant. For the speech presentation mode, speech with easy tracking was better than single-task speech. For each of the three task conditions, icons were better than text. Reaction times for speech are not directly comparable to RTs for visually presented stimuli because speech has duration in time whereas the visual presentation is effectively instantaneous. The speech RTs in this experiment were measured from the onset of each word and so are longer than the time from which the word is apprehended. The average word duration was about 430 ms, so if the RTs had been measured from the end of each word, they would have been 430 ms shorter, which would be faster than either of the visual conditions. However, words are often recognized before they end [22], so this could be an underestimate. The most interesting thing about the RT for speech is that it did not increase under dual-task conditions.

The percentage of errors was based only on items to which the subjects actually responded, and missed responses were not included. To compensate for nonhomogeneity of variance, the two-way repeated measures analysis of variance was performed by using arcsin transformed scores. Presentation mode had a significant effect, $F(2,164) = 51.9, p > 0.001$, as did task condition, $F(2, 164) = 16.5, p > 0.001$, and there was a significant interaction, $F(4, 328) = 4.49, p < 0.01$.

Multiple comparison tests on the percent error data were carried out using the Tukey HSD test [21] ($p < 0.01$). For the text presentation mode but not for icons or speech, single-task performance was significantly better than both of the dual-task conditions, and the differences between dual task with easy tracking and dual task with hard tracking were not significant. Both of the dual-task conditions for text and all three of the speech conditions had significantly more errors than the corresponding conditions for the icon presentation mode. On the whole, the pattern of results for errors on the classification task was reasonably similar to the RT results, but a speed-accuracy tradeoff could not be entirely ruled out. The average correlation between RT and errors was -0.145 for the single-task conditions and -0.171 for the dual-task conditions.

Individual Differences

In evaluating individual differences, tracking scores and classification RTs were used. The error results were not used because in some of the conditions the error rates were so low that many of the subjects had no errors, and therefore errors were not a sensitive indicator of individual differences. RTs had more opportunity to vary with individual differences, and the overall patterns of RT results and errors were quite similar. Table 4 correlates the pretest scores with tracking performance and with classification RT. Spatial ability was moderately correlated with tracking performance for both single-task conditions and for all dual-task conditions, but spatial ability was not significantly correlated with classification RT, regardless of presentation mode or task condition. No consistent pattern of significant correlations of verbal ability with performance was evident on either task.

Table 4 — Correlations (Pearson's r) of pretest scores with tracking performance and with reaction time on the classification task

	Spatial Ability	Verbal Ability
Task Condition		
Tracking		
Easy	0.356**	0.143
Hard	0.382**	0.116
Easy w/ Icons	0.490**	0.267
Hard w/ Icons	0.456**	0.179
Easy w/ Text	0.521**	0.270
Hard w/ Text	0.540**	0.248
Easy w/ Speech	0.454**	0.290*
Hard w/ Speech	0.432**	0.218
Classification RT		
Icons, Single	-0.232	-0.147
Icons w/ Easy	-0.181	-0.219
Icons w/ Hard	-0.214	-0.180
Text, Single	-0.129	-0.102
Text w/ Easy	-0.065	-0.215
Text w/ Hard	0.038	-0.137
Speech, Single	-0.171	-0.308*
Speech w/ Easy	-0.247	-0.242
Speech w/ Hard	-0.140	-0.178

* $p < 0.01$
** $p < 0.001$

Dual-task performance was generally highly predictable from single-task performance, especially for the tracking task, as shown by the correlations in Table 5. The third column of the table shows the correlation of the residuals (i.e., the variability not explained by single-task performance) from Sessions 1 and 2. The extent to which the residuals are correlated indicates the internal consistency of dual-task performance within individual subjects on each of the two tasks. A high correlation of residuals indicates that individual differences in dual-task performance exist that are not explained by single-task performance (i.e., the skill level on each task) and suggests differences in the ability to time share between tasks or differences in the amount of interference between the two tasks [12, 23].

Clear-cut individual differences existed among subjects in performance on the two tasks, but individual subjects tended to be highly consistent in their performance, regardless of presentation mode. Table 6 shows the intercorrelations of tracking scores and of classification RT across the different presentation modes. It can be seen from the very high correlations that performance on the tracking task was very consistent regardless of the form of the classification task with which it was combined. Classification RT also showed considerable consistency across presentation modes, although the correlations were somewhat lower.

EXPERIMENTS 2 AND 3

We expected single-task performance to exceed dual-task performance for all presentation modes, but on the classification task the speech presentation mode showed no decrement in performance when combined with the tracking task, and there was even significant improvement when speech was combined with easy tracking. It is unlikely that this was due to an unusual tradeoff strategy between the two tasks, since the scores on the tracking task did not show losses that were much

Table 5 — Correlations between single-task and dual-task performance for Session 1 and for Session 2 and the correlation of the residuals for Session 1 with the residuals for Session 2

	Session 1	Session 2	Residuals
Task Condition			
Tracking			
Easy w/ Icons	0.790**	0.771**	0.342*
Hard w/ Icons	0.839**	0.851**	0.336*
Easy w/ Text	0.776**	0.759**	0.462**
Hard w/ Text	0.801**	0.890**	0.500**
Easy w/ Speech	0.756**	0.774**	0.406**
Hard w/ Speech	0.777**	0.876**	0.277
Classification RT			
Icons w/ Easy	0.598**	0.560**	0.643**
Icons w/ Hard	0.415**	0.538**	0.490**
Text w/ Easy	0.492**	0.495**	0.523**
Text w/ Hard	0.368**	0.315*	0.441**
Speech w/ Easy	0.627**	0.703**	0.551**
Speech w/ Hard	0.662**	0.686**	0.523**

* $p < 0.01$
** $p < 0.001$

Table 6 — Intercorrelations of dual-task tracking and reaction time scores across presentation modes

Task Condition			
Tracking Easy			
	Icons	Text	Speech
Icons	1.00		
Text	0.869**	1.00	
Speech	0.837**	0.869**	1.00
Tracking Hard			
	Icons	Text	Speech
Icons	1.00		
Text	0.917**	1.00	
Speech	0.916**	0.906**	1.00
Classification RT with Easy			
	Icons	Text	Speech
Icons	1.00		
Text	0.734**	1.00	
Speech	0.652**	0.705**	1.00
Classification RT with Hard			
	Icons	Text	Speech
Icons	1.00		
Text	0.399**	1.00	
Speech	0.670**	0.584**	1.00

* $p < 0.01$
** $p < 0.001$

larger than those obtained with the other presentation modes, as should be the case if the subjects were attending more to the classification task in this condition. One might also have expected that speech would interfere somewhat less with tracking than either of the visual presentation modes, because of the difficulty of looking at two things at the same time. Tracking performance was better with speech than with text, but tracking performance with icons was not significantly better than speech for easy tracking and significantly better for hard tracking. There are several possible explanations for these results. It may be that the icon presentation mode actually was better than speech under dual-task conditions. This might be the case if there is a cost associated with switching between visual and auditory modes as suggested by Wickens and Liu [24] or if the auditory stimuli tended to preempt attention, thereby distracting from the tracking task. Another possibility is that the synthesized speech stimuli required more effort because they were difficult to understand and therefore distracted more from the tracking task.

Experiment 2 was conducted to determine whether the effects of the speech presentation mode in Experiment 1 were primarily due to using synthesized speech that was hard to understand or whether these effects would be the same even when highly intelligible human speech was used. There were three speech conditions in Experiment 2, natural human speech, a high quality commercial synthesizer (DECtalk), and the developmental synthesizer that was used in the first experiment.

Experiment 3 compared the natural and synthesized speech used in Experiment 2 directly with icons and text. In addition, the scoring method for easy and hard tracking was equated so that performance on the two conditions could be directly compared.

Subjects

In Experiment 2, 23 of the original subjects from the NRL group in Experiment 1 were retested. In Experiment 3, 15 University of Maryland undergraduate psychology students volunteered to participate for extra course credit.

Method

The single- and dual-task procedures were similar to those used for Experiment 1. Table 7 shows the design for Experiment 2. Verbal and spatial abilities were not retested. A short practice period was used to familiarize the subjects with the three speech types and to refresh single-task and dual-task skills because several months had passed since Experiment 1. Each of the three speech conditions was then tested singly and in combination with hard tracking. Easy tracking was not included because of time constraints.

The speech stimuli for the developmental synthesizer and for the DECtalk synthesizer were recorded at the rate of one word every 2.5 s in the same way as for Experiment 1. For human speech, a male speaker was recorded reading the lists. The timing was controlled by having the speaker wear headphones over which he heard a tone every 2.5 s, and he was instructed to read the words in synchrony with the tones.

The design for Experiment 3 was similar to that for Experiment 1, except that there were two speech conditions (human and DECtalk) in addition to the icon and text conditions. Table 8 shows the test conditions and presentation orders for the four counterbalanced groups. The scoring method for the easy and hard tracking conditions was also equated. The visible size of the targets for the easy and hard versions were the same as in the previous experiments so that the task appeared the same to the subjects, but the invisible target area on which scores were based was a 30 by 30 pixel square for both conditions. For Experiment 3, the error in the RT program for the speech conditions was also corrected.

Table 7 — Test conditions for Experiment 2

GROUP 1	GROUP 2
SESSION 5	
Single-Task Classification	
Human DECtalk NRL Synthesizer	Human DECtalk NRL Synthesizer
Dual-Task Classification and Tracking	
DECtalk w/hard NRL Synthesizer w/hard	DECtalk w/hard NRL Synthesizer w/hard
Rest Period	
Single-Task Classification Human DECtalk NRL Synthesizer	*Dual-Task Classification and Tracking* Human w/hard NRL Synthesizer w/hard DECtalk w/hard
Single-Task Tracking Hard	*Single-Task Tracking* Hard
Dual-Task Classification and Tracking Human w/hard NRL Synthesizer w/hard DECtalk w/hard	*Single-Task Classification* Human DECtalk NRL Synthesizer

Results

The data were analyzed as for Experiment 1. Figure 3 shows the results for Experiment 2. Repeated measures analysis of variance showed a significant effect for the tracking task, $F(3,66) = 35.7$, $p = < 0.001$. Multiple comparison tests ($p < 0.01$) showed that single-task tracking performance was significantly better than dual-task performance for all three speech types on the classification task, but there were no significant differences among the dual-task tracking conditions regardless of speech type. Reaction times were estimated from the slow response ratio as in Experiment 1. A two-way repeated measures analysis of variance using log-transformed scores showed a significant effect of speech type, $F(2,40) = 13.7$, $p < 0.001$; of task condition $F(1,20) = 34.1$, $p < 0.001$; and a significant interaction $F(2,40) = 52.3$, $p < 0.001$. Multiple comparison tests ($p < 0.01$) showed that single-task performance was significantly better than dual-task performance for human speech and for the DECtalk synthesizer, but there was no difference between single- and dual-task performance for the poorer developmental synthesizer. Single-task performance was significantly better for human speech than for either of the synthesizers, and DECtalk was better than the developmental synthesizer, but there were no significant differences among the speech types on dual-task performance.

Table 8 — Test conditions for Experiment 3

GROUP 1	GROUP 2	GROUP 3	GROUP 4
		SESSION 2	
		Single-Task Classification	
Icon	Text	DECtalk	[See Groups 1,2,3]
Text	DECtalk	Icon	
DECtalk	Icon	Text	
		Single-Task Tracking	
Easy	Easy	Easy	Easy
Hard	Hard	Hard	Hard
		Rest Period	
		Dual-Task Classification and Tracking	
Text w/easy	DECtalk w/easy	Icon w/easy	[See Groups 1,2,3]
Text w/hard	DECtalk w/hard	Icon w/hard	
DECtalk w/easy	Icon w/easy	Text w/easy	
DECtalk w/hard	Icon w/hard	Text w/hard	
Icon w/easy	Text w/easy	DECtalk w/easy	
Icon w/hard	Text w/hard	DECtalk w/hard	
		SESSION 3	
		Single-Task Classification	
Icon	Text	DECtalk	Human
Text	Icon	Human	DECtalk
Human	DECtalk	Text	Icon
DECtalk	Human	Icon	Text
		Rest Period	
		Dual-Task Classification and Tracking	
Human w/easy	DECtalk w/easy	Icon w/easy	Text w/easy
Human w/hard	DECtalk w/hard	Icon w/hard	Text w/hard
DECtalk w/easy	Human w/easy	Text w/easy	Icon w/easy
DECtalk w/hard	Human w/hard	Text w/hard	Icon w/hard
Icon w/easy	Text w/easy	Human w/easy	DECtalk w/easy
Icon w/hard	Text w/hard	Human w/hard	DECtalk w/hard
Text w/easy	Icon w/easy	DECtalk w/easy	Human w/easy
Text w/hard	Icon w/easy	DECtalk w/hard	Human w/hard

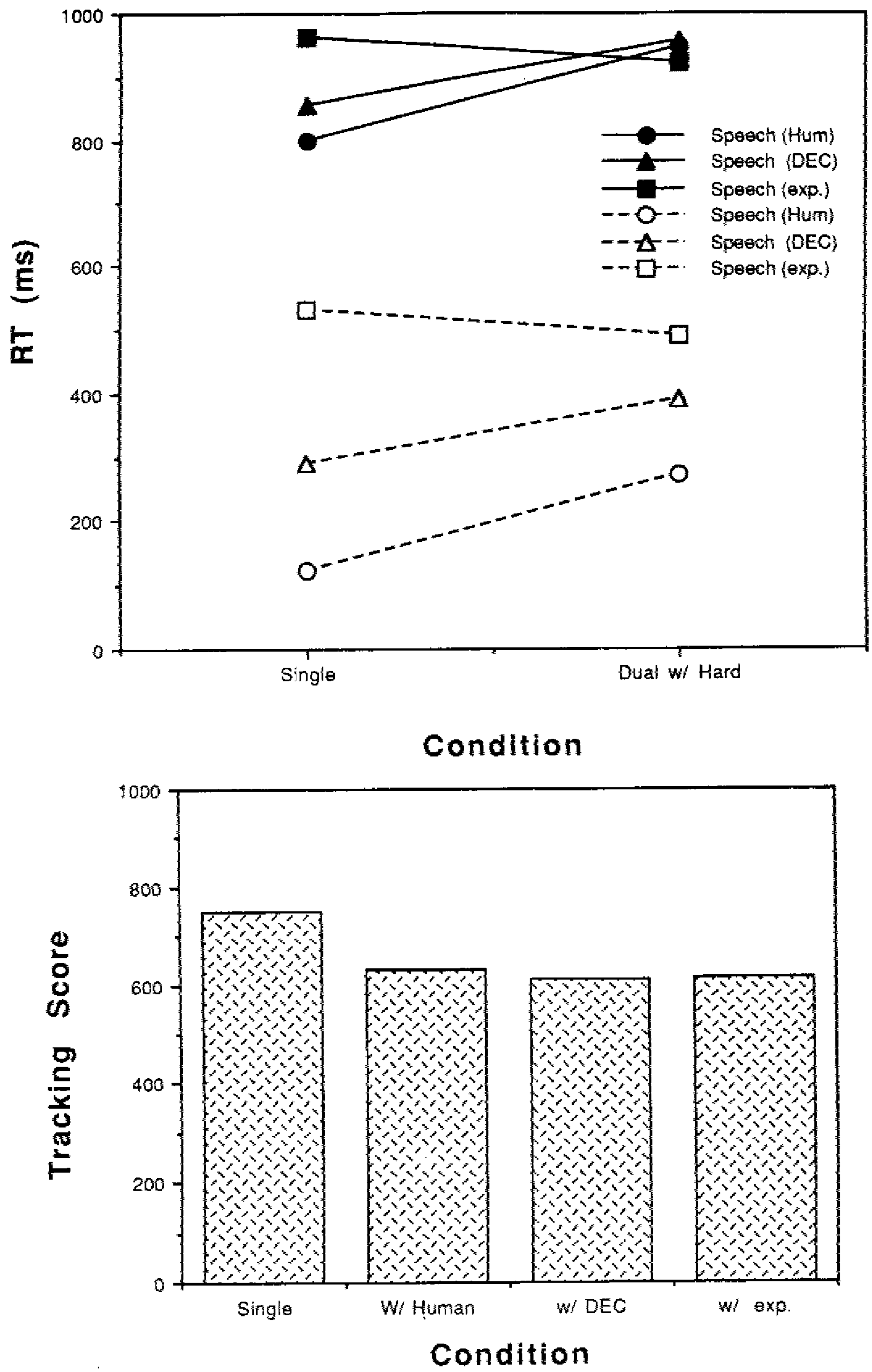

Fig. 3 — Classification and tracking performance for Experiment 2. Classification RT for the speech presentation modes is shown twice, from word onset (solid line) and from word offset (broken line).

Figure 4 shows the results for Experiment 3. Repeated measures analysis of variance for the tracking task showed a significant effect due to presentation mode on the concurrent task, $F(3,42) = 4.14$, $p < 0.02$, but multiple comparison tests failed to show significant pairwise comparisons. Scores for hard tracking were significantly better than for easy tracking, $F(1,14) = 30.33$, $p < 0.001$. Repeated measures analysis of variance for the log-transformed RT scores on the classification showed a significant effect of presentation mode, $F(3,42) = 80.2$, $p < 0.001$. However, multiple comparison tests showed no difference between text and icons or between human speech and DECtalk. The two visual conditions were faster than the two speech conditions because speech extends over time, and if the speech RTs had been measured from speech offset instead of from speech onset, the RTs would have been faster for the speech conditions. The effect of task condition (single vs dual with easy or hard tracking) was marginally significant, $F(2, 28) = 3.9$, $p < 0.05$; and there was a significant interaction, $F(6,84) = 20.9$, $p < 0.001$. Multiple comparison tests ($p < 0.01$) showed that single-task performance was significantly better than dual-task performance for both text and icons, but not for speech, and the differences between dual task with easy tracking and dual task with hard tracking were not significant.

DISCUSSION AND CONCLUSIONS

Presentation Mode Effects

Performance on both easy and hard tracking was reduced when tracking was combined with classification. The decrement was the least when the classification presentation mode was icons and the greatest for the text presentation mode. Speech interfered more than the visually presented pictures, which would seem contrary to the original predictions of multiple-resource theory [15], although a more recent expansion of the theory [16] has acknowledged inconsistencies in predictions for the relationship between task type and mode compatibilities, particularly where speech presentation is compared with pictorial or symbolic presentation. Other research [25] suggests there may be an attention "switching cost" that is greater when auditory stimuli are presented with a visual task than when two tasks are presented within the visual mode. Yet text interfered more than icons or speech. Reading words seems to distract more from tracking than either hearing words or seeing pictures.

The comparison of RTs for speech and visual stimuli is not straightforward. Speech extends over time and is not complete until the entire word has been spoken. Pictures and words, on the other hand, extend over space, and the entire stimulus is present as soon as it is displayed. In experiments with speech stimuli, RT is generally measured from the offset rather than the onset of the speech stimulus. When measured from stimulus onset, classification was faster for both visual modes, icon and text, than for speech, but it was faster for speech if RT was measured from the end of each word. Spoken words may be identified before the end of the word is reached but obviously cannot be recognized until some time after the onset. Likewise, words and pictures are not recognized the instant they appear, but all of the information is available immediately. Even though it is difficult to compare single-task classification performance for speech and visual modes, it can still be noted that in this task, the icon mode was superior to the text mode.

Dual-task classification was slower than single-task classification for both of the visual modes; also, hard tracking interfered more with the classification task than did easy tracking. With the speech mode, however, there was no loss in RT (and possibly an improvement in RT) from single- to dual-task performance. Wickens and Liu [24] describe "preemption" as being similar to switching, but with the additional influence of the 'alerting' nature of auditory displays, relative to their visual counterparts. The consequences of preemption to performance are in favor of the auditory task when combined with a visual task, however. They suggest that discrete auditory stimuli presented concurrently with an ongoing visual task would be likely to draw attention to themselves, thereby diverting attention from the visual task. This interpretation seems to be in agreement with the fact that the speech interfered more with tracking than did icons. The synthesized speech in Experiment 1 was difficult to understand, and therefore subjects may have taken longer to decide because they were unsure of the words. Experiments 2 and 3 demonstrated

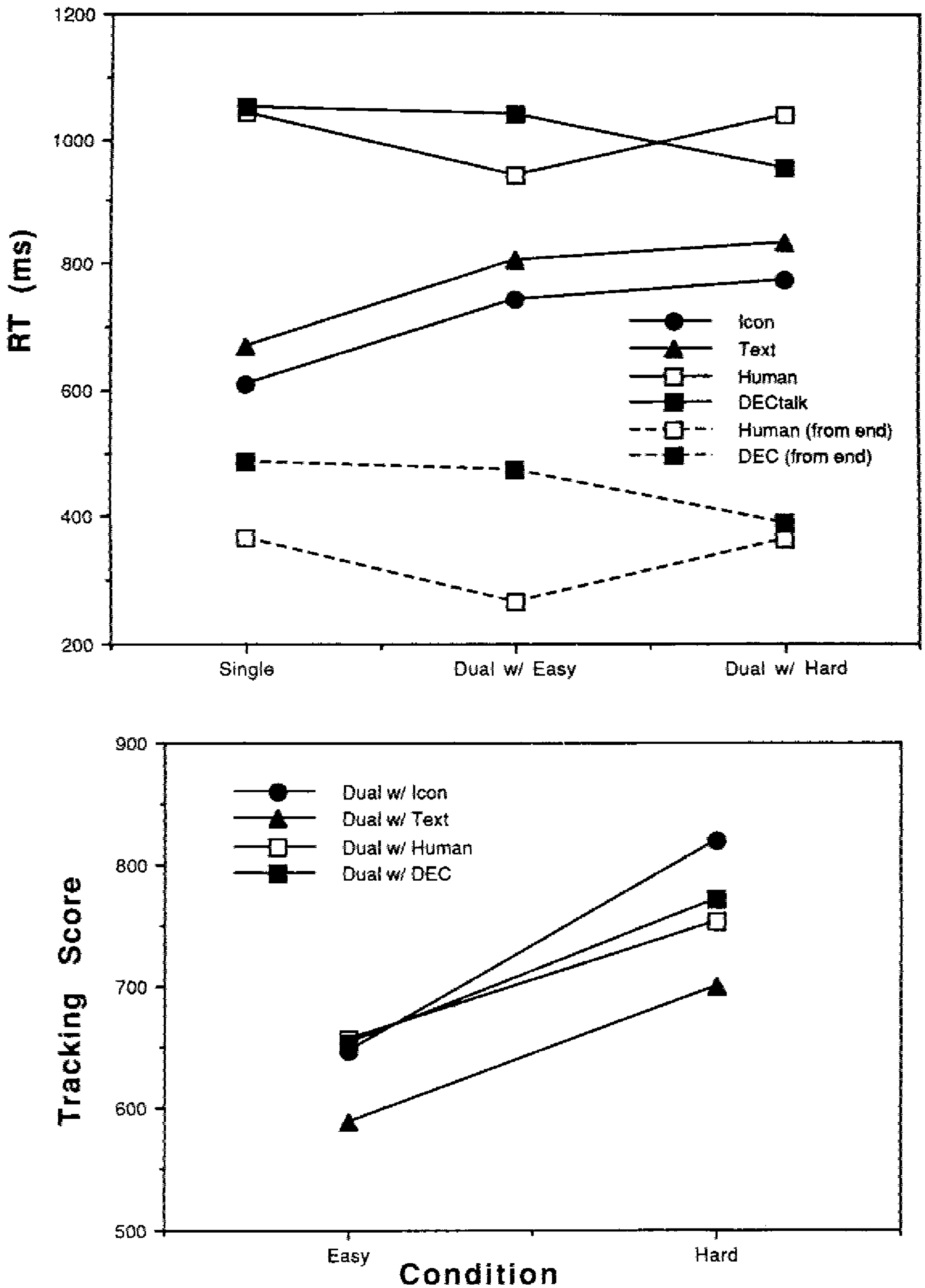

Fig. 4 — Classification and tracking performance for Experiment 3. To facilitate comparisons with the icon and text modes, classification RT for the speech presentation mode is shown twice, from word onset (solid line) and from word offset (broken line).

that single-task RT was faster for human speech and good synthesis than for the experimental synthesis, but the results of all three experiments taken together suggest there was not a dual-task decrement for synthesized speech in combination with tracking and possibly a slight decrement for human speech.

An issue that was not specifically addressed in these experiments is the distinction between data limited processing and resource limited processing [25]. Data limited processing occurs when not enough data is available in the situation to perform a task. Increasing the amount of effort or resources applied to the task will not improve performance because the information needed to perform the task is simply not available. Examples of data limited performance might occur with a blurred or indistinct visual display or with very noisy and unintelligible speech. Resource limited processing occurs when information is coming so fast or from so many sources that it cannot all be attended to at the same time. The information needed to perform the task is available, but the attentional or resource demands on the individual are so great that it is not possible to do the task well or at all.

The effects of data and resource limitations were confounded in the first experiment in that the the quality of the synthesized speech limited the subjects' ability to understand the words while at the same time the dual-task condition involved high resource demands. The text and icon presentation modes, on the other hand, had no such data limitations. The results of the second experiment suggest that the addition of a concurrent tracking task does affect the time required to understand normal highly intelligible speech even though it did not add to the time it takes to process the already difficult speech stimuli. The effects of different types of resource demands were only partially separated in the tasks that were used in these experiments and need to be better distinguished in any follow up investigations.

Individual Differences

As is to be expected, large individual differences in performance were exhibited on the experimental tasks. There were also large individual differences in verbal and spatial abilities as measured by the pretests. Spatial ability was statistically significantly but moderately correlated with performance on the tracking task both for single-task and in combination with the various versions of the classification task. However, verbal ability was not consistently correlated with performance on any of the tasks used in this experiment. This is not surprising with respect to the tracking task , but some correlation with performance on the text or speech presentation modes for the classification task might have been expected, given the similarity of the Goldberg et al. [17] task to our classification task. It may be that because the items to be classified were familiar objects, performance of the task relied more on accessing semantic knowledge than on specific verbal abilities.

Performance consistency between single- and dual-task can be examined by considering the extent to which dual-task performance can be predicted from single-task performance, the residuals for session one and session two were highly intercorrelated. This indicates that the individual subjects were very consistent with themselves on dual-task performance even after taking into account their skill on each of the separate tasks as indicated by single-task performance. That is, there were consistent individual differences in the ability of subjects to combine the two tasks as well as in their ability to perform the two tasks separately. This result is consistent with the findings of Yee et al. [13] although their procedure involved coordinating tasks, whereas the present procedure involved competing tasks.

Even though there were consistent individual differences in the ability to perform and combine the two tasks, there was no indication of changing preferences or strategies depending on the presentation mode for the classification task. As indicated by the high correlations across presentation modes, the performance of individual subjects remained consistent on both classification and tracking performance. Classifying subjects into good and poor performers on each task showed that some subjects performed better on one task than on the other, but there was no evidence that tradeoffs between the two tasks differed for the icon, text, or speech presentation modes. Across experimental conditions, good performers on a particular task tended to stay good on that task, and poor performers tended to stay poor. The responses to subjective questions about strategies showed no evidence of being related to performance tradeoffs. The lack of evidence for individual differences in changing tradeoff strategies with different presentation modes suggests that the presentation mode for a given task can be selected on the basis of the best overall performance rather than being adapted to individual needs. This is an encouraging result in that it means that the presentation format for displaying a given type of information does not need to be adapted to the individual but can be selected on the basis of best overall performance. In the light of resource limitations, it is reasonable to suppose that the effects of individual differences in ability will be more apparent when capacity is strained than when resource demands are low.

ACKNOWLEDGMENTS

We wish to thank all of the friends and colleagues who have helped us with this experiment. Special thanks go to Brian Potter and Tim Wicinski for their work on writing the programs, to Steve Monaco for help with generating additional items for the paper folding test, to Don Kallgren for lending us his voice, and to Stephanie Everett for help with establishing the timing tone set and for lending her synthesizer to produce the word set. We are especially grateful to Susan Feldman who helped extensively in testing subjects and with seeking out references and literature reviews. Finally, our colleagues at NRL who volunteered several hours of their time to participate in the experiment deserve a big thank you. We also thank Jerry Owens who read and commented on an earlier version of this report.

REFERENCES

1. E. Hunt, "What Does It Mean to be High Verbal? " *Cognitive Psychology* **7**, 194-227 (1975)

2. E. Hunt, "Mechanics of Verbal Ability," *Psychological Review* **85**, 109-130 (1978).

3. E. Hunt, "On the Nature of Intelligence," *Science* **219**, 141-146 (1983).

4. E. Hunt, J. W. Pellegrino, R. Frick, S. A. Farr, and D. Alderton, "The Ability to Reason about Movement in the Visual Field," *Intelligence* **12**, 77-100 (1988).

5. J. W. Pellegrino, E. B. Hunt, R. Abate, and S. Farr, "A Computer-Based Test Battery for the Assessment of Static and Dynamic Spatial Reasoning Abilities," *Behavior Research Methods, Instruments, & Computers* **19**, 231-236 (1987)

6. D. E. Egan and L.M. Gomez, "Assaying, Isolating, and Accommodating Individual Differences in Learning a Complex Skill," *Individual Differences in Cognition* **2**, 173-217 (1985).

7. R. D. Peters, G. T. Yastrop, and D. A. Boehm-Davis, "Predicting Information Retrieval Performance," Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA, Human Factors Society, 1988, pp.301-305.

8. R. B. Ekstrom, J. W. French, and H. H. Harmon, "Cognitive Factors: Their Identification and Replication, " *Multivariate Behavioral Research Monographs* **79**(2), (1979).

9. C. D. Wickens, S. J. Mountford, and W. Schreiner, "Multiple Resources, Task-Hemispheric Integrity, and Individual Differences in Time Sharing," *Human Factors* **23**, 211-229 (1981).

10. D. L. Damos, T. E. Smist, and A. C. Bittner, Jr., "Individual Differences in Multiple-Task Performance as a Function of Response Strategy, " *Human Factors* **25**, 215-226 (1983).

11. J. A. Forrester, "An Assessment of Variable Format Information Presentation," Proceedings of the Aerospace Medical Panel Symposium, Toronto, Canada, 1986 (pp. 9.1-9.13).

12. P. L. Ackerman, W. Schneider, and C. D. Wickens, "Deciding the Existence of a Time-Sharing Ability: A Combined Methodological and Theoretical Approach," *Human Factors* **26**, 71-82 (1984).

13. P.L.Yee, E. Hunt, and J. W. Pellegrino, "Individual Differences in the Ability to Integrate Information from Multiple Sources," University of Washington, Department of Psychology, Seattle, 1988.

14. C. D. Wickens, M. Vidulich, and D. Sandry-Garza, "Principles of S-C-R Compatibility with Spatial and Verbal Tasks: The Role of Display-Control Location and Voice-Interactive Display-Control Interfacing," *Human Factors* **26**, 533-543 (1984).

15. C. D. Wickens, "The Structure of Attentional Resources," in *Attention and Performance VIII*, R. S. Nickerson, ed. (Erlbaum, Hillsdale, NJ, 1980), pp. 239-257.

16. C. D. Wickens, D. L. Sandry, and M. Vidulich,"Compatibility and Resource Competition between Modalities of Input, Central Processing, and Output," *Human Factors* **25**, 227-248 (1983).

17. R. A. Goldberg, S. Schwartz, and M. Stewart, "Individual Differences in Cognitive Processes," *Journal of Educational Psychology* **69**, 9-14 (1977).

18. J. G. Snodgrass and M. Vanderwart, "A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity," *Journal of Experimental Psychology: Human Learning & Memory* **6**, 174-215 (1980).

19. J. G. Snodgrass, B. Smith, K. Feenan, and J. Corwin, "Fragmenting Pictures on the Apple Macintosh Computer for Experimental and Clinical Applications," *Behavior Research Methods, Instruments, & Computers* **19**, 270-274 (1987).

20. D. F. Lohman, "Spatial Ability: A Review and Reanalysis of the Correlational Literature," Tech. Rep. 8, Stanford University, School of Education, Stanford, California, 1979.

21. B. J. Winer, *Statistical Principles in Experimental Design*, 2nd ed. (McGraw-Hill, New York, 1971).

22. C. B. Mills, "Effects of Match Between Listener Expectations and Coarticulatory Cues on the Perception of Speech," *Journal of Experimental Psychology: Human Perception and Performance* **6**, 528-535 (1980).

23. E. B. Hunt, J. W. Pellegrino, and P. L. Yee, "Individual Differences in Attention," *The Psychology of Learning and Motivation* **24**, 285-310 (1989).

24. C. D.Wickens and Y. Liu, "Codes and Modalities in Multiple Resources: A Success and a Qualification," *Human Factors* **30**, 599-616 (1988).

25. D. LaBerge, P. VanGelder, and S. Yellott, "A Cueing Technique in Choice Reaction Time," *Journal of Experimental Psychology* **87**, 225-228 (1971).

26. D. Norman and D. Bobrow, "On Data-Limited and Resource-Limited Processes," *Cognitive Psychology* **7**, 44-64 (1975).