

# Dual-task performance using degraded speech in a sentence-verification task

ASTRID SCHMIDT-NIELSEN, HOWARD J. KALLMAN, and CORINNE MEIJER  
*Naval Research Laboratory, Washington, D. C.*

The results of a preliminary study on the effects of reduced speech intelligibility on dual-task performance are reported. The speech task was a sentence-verification task, and speech degradation was accomplished using a narrowband digital voice-transmission system operating with and without random bit errors. The second task was a visual picture-sorting task. There was a dual-task decrement on the sorting task and, in addition, there was a further decrease in sorts per minute as the speech was increasingly degraded. Reaction time for the speech task increased with the concurrent sorting task, but the dual-task condition did not affect speech-task error rates.

There are many voice-communication situations in which the communicator is performing some other task in addition to listening to and understanding the speech—for example, flying an airplane or driving a vehicle. Often these situations are also circumstances in which speech intelligibility can be expected to be more or less degraded due to the transmission channel or background noise. Small losses in intelligibility have little effect on whether ordinary speech is comprehended, and with the more distinctive and highly constrained vocabularies used in military speech, more degradation can be tolerated. However, even if the message is correctly understood, greater mental effort may be required by the listener to understand the speech. With very severe degradations, not only is listener effort further increased, but errors in comprehension occur. Most research on the effects of speech degradation on intelligibility and other performance measures, such as reaction time (RT), has been conducted with the speech task as the only task. In this paper, we report the results of a preliminary study on the effects of reduced speech intelligibility on dual-task performance.

The ability to perform two tasks simultaneously depends on the amount and efficiency of the mental resources allocated to the tasks (e.g., Navon & Gopher, 1979; Norman & Bobrow, 1975). According to multiple resource theory (e.g., Wickens, 1984), the degree to which two tasks will interfere with one another depends on both the resource demands and the resource composition of the tasks. If the tasks draw on different resources, this should result in less interference than if they draw on the same resources, although recent evidence (Wickens & Liu, 1988) suggests that there may also be a cost associated with scanning between different modalities. Regardless

of which version of resource theory we accept, we usually expect some decrement in performance when two tasks are performed simultaneously, and it is what happens when the difficulty of the speech task is increased that is of particular interest here.

We selected a visual picture-sorting task as the second task. In many real situations in which degraded voice communication occurs, the concurrent task demands are likely to be visual-spatial in nature. According to multiple resource theory, less competition can be expected between a visual and a verbal task than between two verbal tasks. It is probable that some overall decrement would occur for both tasks because of dual-task demands. If the mental resources required to process the speech task are separate from those needed for the visual task, the more difficult degraded speech conditions should affect performance only on the speech task. If there is also an additional decrement on the visual task due to increasing the speech difficulty, this would suggest that both tasks are drawing on shared resources and that the extra effort required for the speech task interferes with the processing of the visual task as well.

The speech task we used was a sentence-verification task (Larkey & Danly, 1983), and speech degradation was accomplished using a narrowband digital voice-transmission system operating with and without random bit errors. Narrowband digital voice-transmission systems for secure voice applications, both commercial and military, are coming into increasingly widespread use. A linear predictive coding (LPC) algorithm operating at 2.4 kbps has been established as the Department of Defense standard (MIL-STD-199-113 or Federal Standard 1015). Enhanced versions of this algorithm (LPC-10e) have been incorporated in the Subscriber Terminal Unit (STU-III) and in the Navy's Advanced Narrowband Digital Voice Terminal. Even slight losses in quality can lower the scores on intelligibility tests such as the Diagnostic Rhyme Test (DRT) (Voiers, 1977), which measures the discriminability of word pairs differing only in a single distinctive feature. Intelligibility tests indicate that

---

This research was conducted while Howard J. Kallman held an Office of Naval Technology postdoctoral fellowship. Corinne Meijer is now at the Department of Psychology, University of Maryland, College Park, MD 20742. Correspondence may be addressed to Astrid Schmidt-Nielsen, Human Computer Interaction Laboratory, Naval Research Laboratory, Code 5532, Washington, D.C. 20375-5000.

although scores for LPC-10e processed speech are lower than those for wideband speech, intelligibility is nevertheless quite good, with a score of 89.8 on the DRT (Sandy, in press) and 98% correct recognition of the words of the International Civil Aviation Organization spelling alphabet and digits (Schmidt-Nielsen, 1987). Scores for undegraded speech were 97.6 for the DRT and 99% correct for the spelling alphabet. For LPC with 5% random bit errors, the DRT score falls to about 75 and only about 90% of the spelling alphabet words are correctly understood. Kallman and Schmidt-Nielsen (1989) used a sentence-verification task to evaluate the comprehension of unprocessed speech and of LPC-processed speech with and without random bit errors. Response accuracy decreased and RT increased with LPC processing and with increasing bit error rate.

## METHOD

### Tasks

The two experimental tasks were a sentence-verification task, in which listeners decided whether a spoken sentence was true or false, and a timed card-sorting task. The listeners were seated at a table and had a response box with two pushbuttons, labeled TRUE and FALSE, by the preferred hand. A master card with pictures of 24 common objects was placed by the nonpreferred hand, with an oblong box beside it containing small cards with pictures of individual objects. For the sentence-verification task, the listeners heard 28 sentences over high-quality headphones and used the index and middle fingers of the preferred hand to press the labeled pushbuttons. They were to decide whether each sentence was true or false and to push the appropriate button as quickly as possible while avoiding mistakes. The card-sorting task was timed to last barely longer than the sentence presentation, so that in the dual-task condition, sorting began just before the first sentence and ended just after the last one. The words "start" and "stop" spoken in a female voice and heard over the headphones were used to cue the subjects when to begin and end the sorting task. The subjects used the nonpreferred hand to pick up the object cards one at a time from the box, turn them over, and place them on the corresponding pictures on the master card. The object was to sort as many cards as possible during the time interval between the words "start" and "stop."

### Speech Materials

The speech materials were the same as those used by Kallman and Schmidt-Nielsen (1989). The test items consisted of 96 true and 96 false sentences. Sixty additional sentences were used for practice and for fillers. The practice list and the eight test lists had 28 items each. The first four items (two true and two false) in each test list were fillers and were not scored. The remaining 24 items in each list were the test sentences consisting of equal numbers of true and false statements equally distributed across strong and weak property and category relationships. The lists were recorded by a male speaker with approximately 2 sec of silence between sentences.

The test tapes were identical to those used by Kallman and Schmidt-Nielsen (1989). Four speech conditions were tested: high-quality unprocessed speech and LPC-processed speech with 0%, 2%, and 5% random bit errors. Four counterbalanced sequences of eight lists each were prepared. The order of the four processing conditions was balanced across sequences, and the order in which the different processing conditions were presented in the second half of each sequence was the reverse of the order in the first half to control for possible practice and/or fatigue effects.

### Sorting-Task Materials

The stimuli for the card-sorting task were 90 black-and-white drawings of common objects. Arbitrarily selected subsets of 24 pictures were used to generate 15 different master-card layouts (4 rows  $\times$  6 columns),

such that no two master cards had more than six pictures in common. For each master card, a stack of object cards was made with multiple copies of each of the 24 pictures from that master card. A different master card was used for each trial, and the object cards were shuffled before each trial and placed in a small rectangular box next to the subject, with the picture side of the cards facing away from the subject.

### Subjects and Procedure

The listeners were 48 undergraduate psychology students at the University of Maryland (12 for each of the four sequences) who volunteered to participate for extra course credit. The listeners were tested individually. Each listener was first familiarized with the sound of LPC speech by listening to LPC-processed versions of five different speakers, each reading the same 30-sec paragraph. This was followed by a practice trial on the sentence-verification task, in which the subjects heard the practice list in a fairly difficult speech condition, LPC-processed speech with 2% bit errors. After practice on the verification task, the subjects were given two practice trials on the sorting task alone. The final practice trial combined the two tasks, again using LPC-processed speech with 2% bit errors. Each subject then completed one of the counterbalanced sequences of eight dual-task test lists, with a 5-10-min break between the first and second half of testing. At the end of the session, there was a single trial of card sorting alone.

### Scoring Procedure

The responses and RTs to the sentence-verification task were collected and stored by an IBM PC computer. The RTs were calculated from the end of the last word of each sentence as determined by visual inspection of the digitized waveform. The number of cards correctly sorted was divided by the time from start to stop in order to obtain the number of correct sorts per minute for each trial.

## RESULTS

Separate analyses of variance were carried out for each of the dependent measures: RT and percentage of correct responses for the sentence-verification task, and number of pictures sorted per minute for the picture-sorting task. The degrees of freedom for the *F* tests were corrected, where appropriate, for violations of sphericity using the Huynh and Feldt (1976) correction.

The effect of speech-processing condition on the sentence-verification and sorting tasks is shown in Table 1. As expected, RTs and errors for the sentence-verification task increased as the intelligibility of the speech conditions decreased [ $F(2.95, 138.87) = 63.30$ ,  $MS_e = 62,434$ ,  $p < .0001$ , and  $F(2.74, 128.93) = 111.95$ ,  $MS_e = 154.17$ ,  $p < .0001$ , for RTs and errors, respectively]. In addition to affecting the verification task, increasing the difficulty of the speech conditions also led to decreased performance on the picture-sorting task [ $F(2.99, 140.35) = 20.69$ ,  $MS_e = 3.49$ ,  $p < .0001$ ].

**Table 1**  
Effect of Speech-Processing Condition on Mean Reaction Time and Percentage of Errors for the Sentence-Verification Task and on Number of Pictures Correctly Sorted for the Sorting Task

Condition	Verification Task		Card-Sorting Task
	Mean RT	% Errors	Sorts/Minute
Unprocessed	551.0	5.7	23.07
LPC 0% ber	646.5	9.5	22.87
LPC 2% ber	701.9	13.2	22.39
LPC 5% ber	793.3	21.4	21.14

Note—LPC = linear predictive coding. ber = bit error rate.

This suggests that the increased effort required to understand the more difficult speech left less processing capacity available for performing the sorting task.

Figure 1 compares dual-task performance with single-task performance. Single-task performance for the picture-sorting task was determined by averaging the second single-task practice sort before dual-task testing with the final single-task sort after dual-task testing. It can be seen that there was a decrease in overall performance from the single-task to the dual-task condition as well as an addi-

tional decrease due to decreasing speech quality on the verification task. Single-task data on the sentence-verification task were not obtained in this experiment because of time constraints. For comparison purposes, the data from the earlier single-task sentence-verification study by Kallman and Schmidt-Nielsen (1989) are replotted here. When we compared mean RTs and error rates on the practice trials for the two experiments, they were virtually identical. This suggests that the two experiments were reasonably comparable. A comparison of the sentence-verification results of the present experiment with those from the previous experiment suggests that adding a sorting task had no effect on overall error rates but did increase the response time in the sentence-verification task.

The data analyses also showed some effects of practice in the first half versus the second half comparisons. The percentage of verification errors decreased from 13.4% to 11.5% [ $F(1,47) = 8.07, MS_e = 161.99, p < .01$ ] and the number of correct sorts per minute increased from 21.5 to 23.3 [ $F(1,47) = 58.73, MS_e = 4.54, p < .001$ ], but there was no improvement in RTs [ $F(1,47) = 1.20, MS_e = 62.434$ ]. The reduction in the error rate suggests that practice with the LPC-processed speech made it easier to understand, and this finding is in agreement with previous research on LPC speech, which shows rapid improvement with practice over the first few trials (e.g., Schmidt-Nielsen & Everett, 1982). The improvement on the sorting task seems to be largely attributable to a gradual overall increase in performance over trials.

DISCUSSION

Relative to single-task performance, there was a decrement in performance on the card-sorting task in the dual-task condition, and the number of correct sorts decreased further with increasing speech degradation. The fact that performance on the card-sorting task decreased from the single-task to the dual-task condition is not surprising. The additional decrease in performance with increasing difficulty of the speech task suggests that, over and above the expected dual-task decrement, the extra effort required to understand the degraded speech further decreases the mental processing capacity available for performing the sorting task.

The comparison between single-task and dual-task performance for the verification task involves comparing the present results with the results of a previous experiment, so any conclusions should be considered somewhat tentative, even though the two subject groups were very similar in performance. For both single-task and dual-task conditions, the percentage of verification errors increased with the less intelligible speech conditions, as would be expected, but there were no more errors under dual-task conditions than when the verification task was performed alone. The addition of the sorting task did not affect overall comprehension of the sentences. There was, however, an increase in RT under dual-task conditions. It appears that the increase in RT for the sentence-verification task was fairly constant across processing conditions, which is to be expected since the sorting task did not vary in difficulty.

Both tasks have a cognitive component and both require a motor output, so some competition for resources is to be expected and coordination between the motor outputs is required. One can expect some output interference between the two tasks since both the sorting task and the buttonpress for the verification task required a motor response. The act of picking up and putting down cards with one hand is likely to lead to somewhat longer RTs to the button push. The act of pressing the button might also result in slightly fewer pictures being sorted if the mechani-

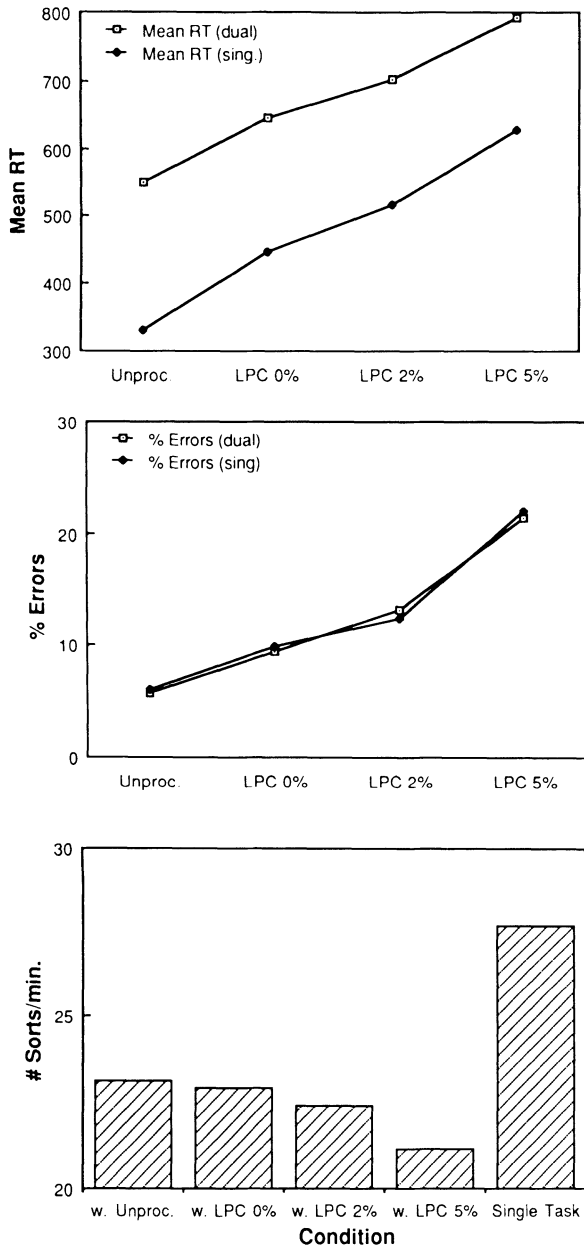


Figure 1. Comparison of mean RT (top panel) and percentage of errors (middle panel) for single-task and dual-task performance for the sentence-verification task, and number of correct sorts per minute (bottom panel) for the picture-sorting task.

cal act of pushing the button with one hand slows down the picking up and putting down of cards with the other hand, but that effect should be small. The mental processing for the sorting task involved identifying the picture and locating the correct match on the layout. The mental processing for the verification task involved comprehension of the sentence and deciding whether the statement was true or false. The additional effort required to understand the speech in the degraded conditions led to a further decrease in card-sorting performance beyond that caused by the dual-task situation alone. Thus, it would appear that both tasks draw on shared resources and that the extra effort required for the speech task interferes with the processing of the visual task as well.

#### REFERENCES

- HUYNH, H., & FELDT, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in the randomized block and splitplot designs. *Journal of Educational Statistics*, *1*, 69-82.
- KALLMAN, H. J., & SCHMIDT-NIELSEN, A. (1989). The effect of LPC narrowband processing and bit error rate on performance in a sentence verification task. *Military Psychology*, *1*, 73-89.
- LARKEY, L. S., & DANLY, M. (1983). Fundamental frequency and sentence comprehension. *MIT Speech Group Working Papers* (Vol. 2).
- NAVON, D., & GOPHER, D. (1979). On the economy of the human processing system. *Psychological Review*, *86*, 214-255.
- NORMAN, D. A., & BOBROW, D. G. (1975). On data limited and resource limited processes. *Cognitive Psychology*, *7*, 44-64.
- SANDY, G. F. (in press). *Digital voice processor consortium report on LPC-10e performance*. McLean, VA: Mitre Corp.
- SCHMIDT-NIELSEN, A. (1987). *Intelligibility of ICAO spelling alphabet words and digits using severely degraded speech communication systems* (Report No. 9035). Washington, DC: Naval Research Laboratory.
- SCHMIDT-NIELSEN, A., & EVERETT, S. S. (1982). A conversational test for comparing voice systems using working two-way communication links. *IEEE Transactions on Acoustics, Speech, & Signal Processing*, *ASSP-30*, 853-863.
- VOIERS, W. D. (1977). Diagnostic evaluation of speech intelligibility. In M. E. Hawley (Ed.), *Speech intelligibility and speaker recognition* (pp. 374-387). Stroudsburg, PA: Dowden, Hutchinson, & Ross.
- WICKENS, C. D. (1984). *Engineering psychology and human performance*. Columbus, OH: Merrill.
- WICKENS, C. D., & LIU, L. (1988). Codes and modalities in multiple resources: A success and qualification. *Human Factors*, *30*, 633-644.

(Manuscript received June 16, 1989.)