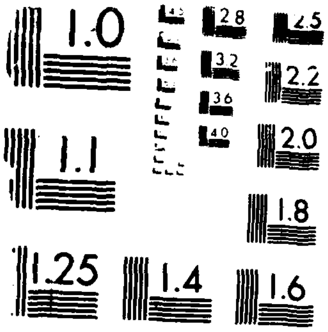
AD-A188 573 EVALUATING THE PERFORMANCE OF THE LPC (LINEAR PREDICTIVE CODING) 24 KBPS ((U) NAVAL RESEARCH LAB WASHINGTON DC A SCHMIDT-NIELSEN ET AL 30 NOV 87 1/1

UNCLASSIFIED NRL-9089 F/G 25/4 NL



MICROCOPY RESOLUTION TEST CHART

4

Naval Research Laboratory
Washington, DC 20375-5000

NRL Report 9089

Evaluating the Performance of the LPC 2.4 kbps Processor with Bit Errors Using a Sentence Verification Task

A. SCHMIDT-NIELSEN AND H. J. KALLMAN

Communication Systems Engineering Branch
Information Technology Division

November 30, 1987

REPORT DOCUMENTATION PAGE	Form Approved OMB No 0704 0188
1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NRL Report 9089	5 MONITORING ORGANIZATION REPORT NUMBER(S)

6a NAME OF PERFORMING ORGANIZATION Naval Research Laboratory	6b OFFICE SYMBOL (If applicable) Code 5526	7a NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code) Washington, DC 20375-5000		7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING SPONSORING ORGANIZATION Office of Naval Research	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code) Arlington, VA 22217		10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N	RR021-05-42		EX 156-044

11 TITLE (Include Security Classification)

Evaluating the Performance of the LPC 2.4 kbps Processor with Bit Errors Using a Sentence Verification Task

12 PERSONAL AUTHOR(S)

Schmidt-Nielsen, A. and Kallman, H. J.

13a TYPE OF REPORT Interim	13b TIME COVERED FROM 10/85 TO 9/86	14 DATE OF REPORT (Year, Month, Day) 1987 November 30	15 PAGE COUNT 18

16 SUPPLEMENTARY NOTATION

(See page ii)

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP	Linear predictive coding	Intelligibility
			LPC	Sentence verification
			Bit errors	Reaction time

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The comprehension of narrowband digital speech with bit errors was tested by using a sentence verification task. The use of predicates that were either strongly or weakly related to the subjects (e.g., A toad has warts./A toad has eyes.) varied the difficulty of the verification task. The test conditions included unprocessed and processed speech using a 2.4 kb/s (kilobits per second) linear predictive coding (LPC) voice processing algorithm with random bit error rates of 0%, 2%, and 5%. In general, response accuracy decreased and reaction time increased with LPC processing and with increasing bit error rates. Weakly related true sentences and strongly related false sentences were more difficult than their counterparts. Interactions between sentence type and speech processing conditions are discussed.

20 DISTRIBUTION AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL Astrid Schmidt-Nielsen	22b TELEPHONE (Include Area Code) (202) 767-2682	22c OFFICE SYMBOL Code 5526

16. SUPPLEMENTARY NOTATION

This research was conducted while Howard J. Kallman held an Office of Naval Technology postdoctoral fellowship. Howard J. Kallman is also at Department of Psychology, State University of New York at Albany, 1400 Washington Avenue, Albany, NY 12222.

CONTENTS

EVALUATING THE PERFORMANCE OF THE LPC 2.4 KBPS PROCESSOR WITH BIT ERRORS USING A SENTENCE VERIFICATION TASK

INTRODUCTION

Digital voice transmission methods are becoming increasingly widespread for ordinary telephone use and for secure voice communications. Some loss in speech quality occurs at the lower data rates required for many secure voice applications. This can affect human performance in various ways depending on the severity of the degradation. Even slight losses in quality can lower the scores on intelligibility tests such as the Diagnostic Rhyme Test (DRT), which measures the discriminability of pairs of words differing only in a single distinctive feature (e.g., *moot-boot* differs only in nasality). Small losses in intelligibility may have little effect on the comprehension of ordinary speech, but greater effort and more time may still be needed for the listener to understand the speech. With more severe degradations, not only is the listener's effort further increased, but errors in comprehension occur. Consequently, in addition to intelligibility tests which measure only errors, it is of interest to investigate methods to assess the time and effort required to comprehend various types of processed speech.

A sentence verification task, in which the listener is required to decide as quickly as possible whether a sentence such as *A giraffe has stripes* is true or false, can be used to evaluate the amount of time necessary to comprehend simple sentences [1]. To the extent that reaction times are long, it can be assumed that greater processing effort is required to comprehend a particular type of sentence or speech processing condition. Manous, Pisoni, Dedina, and Nusbaum [2] demonstrated that reaction times on a sentence verification task were longer for synthetic than for natural speech, even when all of the words were correctly understood. Pisoni and Dedina [3] also used a sentence verification task to evaluate the effect of speech processing and found higher error rates and longer reaction times for 2.4 kilobits per second (kbps) linear predictive coded (LPC) speech than for wideband speech. Longer reaction times that result from poorer quality speech can have negative consequences for performance. For example, in military combat situations where split-second decisions may be required it may take longer to react appropriately to a degraded speech message, even if the message is correctly comprehended.

For narrowband secure voice communications, an LPC algorithm operating at 2.4 kbps has been established as the DoD standard (MIL-STD-199-113 or Federal Standard 1015). Because of the widespread application of this standard, we focused on this type of speech. Versions of this algorithm have been incorporated in the Subscriber Terminal Unit (STU-III) and in the Navy's Advanced Narrowband Digital Voice Terminal (ANDVT) and will consequently be widely deployed. Intelligibility tests indicate that although scores for LPC processed speech are lower than for wideband speech, intelligibility is nevertheless reasonably good, with a score of about 86 on the DRT* and 98% correct recognition of the words of the International Civil Aviation Organization (ICAO) spelling alphabet

---

Manuscript approved July 6, 1987.

*The DRT scores represented in this report are scores obtained using the TRW processor that was used to process the speech samples used in this experiment. This processor employs Version 43 of the DoD standard LPC-10 algorithm. The scores reported by the Digital Voice Processor Consortium [5] are slightly higher, and preliminary results indicate that the new LPC-10e can be expected to score 3 to 5 points higher than the DRT scores reported here.

and digits [4]. High levels of interference or jamming may occur in certain military environments and could result in significant decreases in message intelligibility. One way to simulate a high interference transmission situation is to introduce random bit errors into the transmission stream of the LPC processor. For LPC with 5% random bit errors, the DRT score falls to about 75, and only slightly over 90% of the ICAO spelling alphabet words are correctly understood. Although these results and results obtained by Digital Voice Processor Consortium [5] suggest that transmissions over LPC systems are reasonably comprehensible in the absence of bit errors, and somewhat less so with increasing bit errors, the effect of LPC processing and bit errors on the amount of time that it takes to respond to a message merits investigation. The present experiment was carried out to evaluate the effect of different levels of digital speech degradation on reaction times and comprehension errors in a sentence verification task.

We were also interested in the effect of context on reaction time to and comprehension of processed speech. Military voice communications are generally more robust than ordinary communications because they often employ highly distinctive vocabularies that are designed to be intelligible under adverse conditions. Also, knowledge of the mission context may help to make incoming speech easier to understand, thus accurate communication can be maintained under relatively severe degradations. In other situations, for example normal conversational speech or high level policy discussions, the communication may be more open-ended and fewer contextual constraints would therefore be available to aid comprehension. Knowledge about how contextual information interacts with the effect of speech processing would be useful when evaluating a speech processor for use in a particular environment, because it would make it easier to take into account the degree to which context could be used to aid comprehension. We manipulated context in the sentence verification task by using either strong subject-predicate relationships (e.g., *Camels have humps*) or weak subject-predicate relationships (e.g., *Camels have tongues*) within the sentences.

The context provided by the early part of the sentence can often be used to help disambiguate later words, e.g., Refs. 6 and 7. Thus, in the sentence *Camel have humps*, comprehension of the word *camels* would serve to prime the concept *humps*, because of the strong relationship between the two concepts in semantic memory. Accordingly, perception of the sentence should be facilitated and reaction times to verify the sentence should be shorter. In contrast, *Camels have tongues* expresses a weak subject-predicate relationship, therefore perception of the word *camels* would not be likely to facilitate perception of the word *tongues*. The detrimental effects of LPC processing and bit errors on comprehension should be less for strongly related than for weakly related sentences because the strongly related context should help make the degraded words easier to recognize.

Subject-predicate relatedness should also affect the perception of false sentences, but the overall effect on reaction time should be somewhat different. Although the effect of relatedness may be somewhat smaller for false than for true sentences because the relatedness of the subject and predicate would not be as strong, a relatively highly related context should still help perception more than a weakly related one, because of the priming effect of the earlier words in the sentence on the later words.

In addition to influencing the perception of the words in the sentence, the subject-predicate relatedness variable can also affect decision time, the time it takes to decide whether the sentence is true or false once the words of the sentence have been perceived. Strongly related true sentences express relationships that are more closely associated in semantic memory than weakly related ones and are therefore easier to verify, thus resulting in faster reaction times [8-11]. This effect would probably not be influenced by the difficulty of the speech processing condition because the decision process would occur after the words of the sentence had been perceived. However, for false sentences the decision about whether the content of the sentence is true or false would be more difficult in the strongly related case [9 and 10]. That is, *A fiancee is a relative* would generally be more difficult to reject at the decision stage than *A fiancee is furniture*, since fiancee and relative are associated concepts, whereas fiancee and furniture are not. As with true sentences, the effect of subject-predicate relatedness on the decision stage of processing should remain roughly constant across levels of speech degradation because it is due to decision processes that should be relatively unaffected by the quality of the sensory information. False sentences, however, contrast with true sentences in that strong relatedness has a positive effect on word recognition but a negative effect on the decision stage. Thus, as the quality of the sensory information suffers with increasing degradation of the speech signal, the advantage of weakly related sentences in terms of decision processes would be counterweighed by the advantage of strongly related sentences in terms of perceptual processes, and the advantage of the weakly related false sentences would diminish with LPC processing and with increasing bit errors.

Finally, practice with a particular type of speech processing should result in improved listener performance. The present experiment included a comparison of performance during the first and second halves of testing. Thus, the variables of interest were the speech processing condition, subject-predicate relatedness, and first vs second half of testing. In addition to main effects involving these variables, some interactions of these effects with the truth value of the sentences were predicted.

METHOD

Test Materials

There were 96 true and 96 false sentences, generated so that the subjects and predicates in half of the sentences were strongly related and the subjects and predicates in the other sentences were weakly or not related. The true sentences were generated by drawing on previously published norms and lists of strongly and weakly associated or related property and category relationships, e.g., Refs. 11-14, with additional items that have similar relationships selected and agreed upon by the authors. The false sentences were generated analogously by choosing untrue properties and categories that were either strongly or weakly related to the item in question. For example:

	Strong	Weak
True Property:	A toad has warts.	A toad has eyes.
True Category:	A fly is an insect.	A gnat is an insect.
False Property:	Camels have horns.	Camels have chimneys.
False Category:	Crabs are fish.	Redwoods are fish.

Sixty additional sentences were generated similarly for a practice list and for fillers. The practice list and the eight test lists had 28 items each. The first 4 items (2 true and 2 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 order of the sentences within each list was randomized. The practice list and the test lists were recorded by a male speaker whose voice was known not to create any unusual problems when processed by the LPC algorithm. Approximately 2 s of silence separated consecutive sentences.

Voice Conditions

In addition to high-quality unprocessed speech there were three versions of degraded, LPC-processed speech with 0%, 2%, and 5% random bit errors. The LPC tapes were generated by processing the tape recorded materials through a TRW low data rate voice terminal that uses version 43 of the DoD standard LPC-10 algorithm. For the 2% and 5% bit error conditions, random bit errors were introduced into the bit stream between the analysis and synthesis portions of the processing.

Design

Four counterbalanced sequences of the eight test lists were prepared. Each sequence was divided into halves with one test list for each of the four processing conditions in each half. The order of the processing conditions was balanced across sequences, but the order of the eight sentence lists remained the same across sequences, so that each set of sentences occurred under all four processing conditions. To further balance possible effects of practice or fatigue, 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.

Subjects and Procedure

The listeners were 48 undergraduate psychology students from the University of Maryland, 12 for each of the four sequences, who volunteered to participate for extra course credit. The listeners were tested individually, and the speech was heard through high-quality headphones. Before the sentence verification task, the listeners were familiarized with the sound of LPC speech by listening to LPC-processed versions of five different speakers; each read the same 30-s paragraph. During the experiment, the listeners were seated at a table and placed the index and middle fingers of their preferred hand on two push buttons labeled *true* and *false*. They were told to decide whether each sentence was true or false and to push the appropriate button as quickly as possible without making mistakes. The practice list of 28 sentences, consisting of LPC-processed speech with 2% bit errors, was presented just before the test lists. After the practice, each listener heard one of the sequences of eight test lists, with a 5 to 10 min break between the first and second half of testing.

Scoring Procedure

An IBM PC computer was used to collect and store the responses and reaction times. The reaction times were calculated from the end of the last word of each sentence as determined by visual inspection of the digitized waveform.

RESULTS

Analyses of variance were performed on the reaction time and response error data. Only correct responses were included in the reaction time analysis. In the analyses, the within subjects variables were processing condition, truth value, subject-predicate relatedness, and replication. The degrees of freedom for the F tests were corrected, where appropriate, for violations of sphericity using the Huynh and Feldt correction [15].

As expected, mean reaction time and error rate were greater for LPC than for unprocessed speech and increased progressively with increases in bit error rate. Mean reaction time was 330 ms for the unprocessed speech and 448,516, and 627 ms for LPC speech with 0%, 2%, and 5% bit errors, $F(2.42, 113.77) = 101.24$, $p < .001$, $MSe = 59{,}023$.[*] The corresponding error rates were 6.0%, 9.9%, 12.4%, and 21.9%, respectively, $F(3.29, 137.00) = 137.71$, $p < .001$, $MSe = 127.79$.

When averaged across processing conditions, the main effect of subject-predicate relatedness was significant, and the strongly related sentences had shorter reaction times, $F(1, 47) = 8.90$, $p < .01$, $MSe = 26{,}276$, and fewer errors, $F(1, 47) = 8.57$, $p < .01$, $MSe = 159.66$, than weakly related sentences. There was no advantage of strong relatedness for the unprocessed speech, presumably because strong trues but weak falses have the advantage with respect to decision time. The overall effect is mainly the result of the increasing advantage for the strongly related sentences with increasing degradation, as evidenced by the significant interaction between processor and subject-predicate relatedness for reaction times, $F(2.27, 106.77) = 4.64$, $p < .01$, $MSe = 36{,}819$, and errors, $F(2.12, 99.77) = 7.95$, $p < .001$, $MSe = 170.95$, shown in Fig. 1. In both instances, the effect of processing condition was greater for weakly than for strongly related sentences, presumably because the more strongly related final word was more likely to have been primed or activated by the preceding portion of the sentence, and it would therefore be easier to recognize even when the speech was degraded.

Averaged across conditions, the reaction time to true sentences was faster than to false sentences, with fewer errors for false than for true sentences. Mean reaction times were 404 ms for true and 557 ms for false sentences, $F(1, 47) = 170.42$, $p < .001$, $MSe = 52{,}892$. The respective error rates were 15.5% and 9.6%, $F(1, 47) = 46.71$, $p < .001$, $MSe = 282.30$. At first, these results might appear to suggest a speed-accuracy tradeoff; however, it is more likely that the low error rate for false sentences reflected a bias toward responding *false* when the listener could not understand all of the words, since the proportion of false responses also increased as the speech became more degraded.

There were significant interactions between truth value and processing condition for reaction time and for errors (Fig. 2). The more difficult processing conditions led to greater increases in reaction time for false than for true sentences, $F(2.00, 93.77) = 6.04$, $p < .01$, $MSe = 35{,}936$. If it is inherently more difficult to decide that a sentence is false, then it may be that decreasing the intelligibility of the speech interacts to make this decision even harder. The error rates, on the other hand, increased more for true than for false sentences $F(2.85, 133.77) = 39.63$, $p < .001$, $MSe = 159.32$. If the listeners had a bias to respond *false* when they could not understand a sentence properly, it would have had the effect of depressing the number of correct true responses while inflating the number of correct false responses. Moreover, this effect would be expected to increase as the speech became progressively less intelligible (Fig. 2).

---

[*] A number of statistics are calculated in an analysis of variance. For each main effect or interaction, an F ratio is calculated and forms the basis for determining whether the variable had an effect or interacted with another variable to have an effect on the dependent measure. In general, the higher the F ratio the more likely that the independent variable or variables had an effect on the scores. The F ratio is evaluated with reference to the degrees of freedom of the test, which we have enclosed in parentheses. Although in most cases the reported degrees of freedom are whole numbers, some of our degrees of freedom include fractional values because of our use of a correction for violations of sphericity, a violation often present in repeated-measures designs. Following each report of an obtained F ratio, there is a probability value associated with the particular combination of F ratio and degrees of freedom, a value that can be obtained from commonly available statistical tables. It is conventional among psychologists to assume that if the probability of obtaining a given F ratio by chance alone is less than .05, the effect of the variable or combination of variables on the scores is statistically significant and represents a real effect. Finally, we have reported the mean square errors, which are used in calculating the F ratio and are measures of the amount of random variability in the scores that underlie each test.

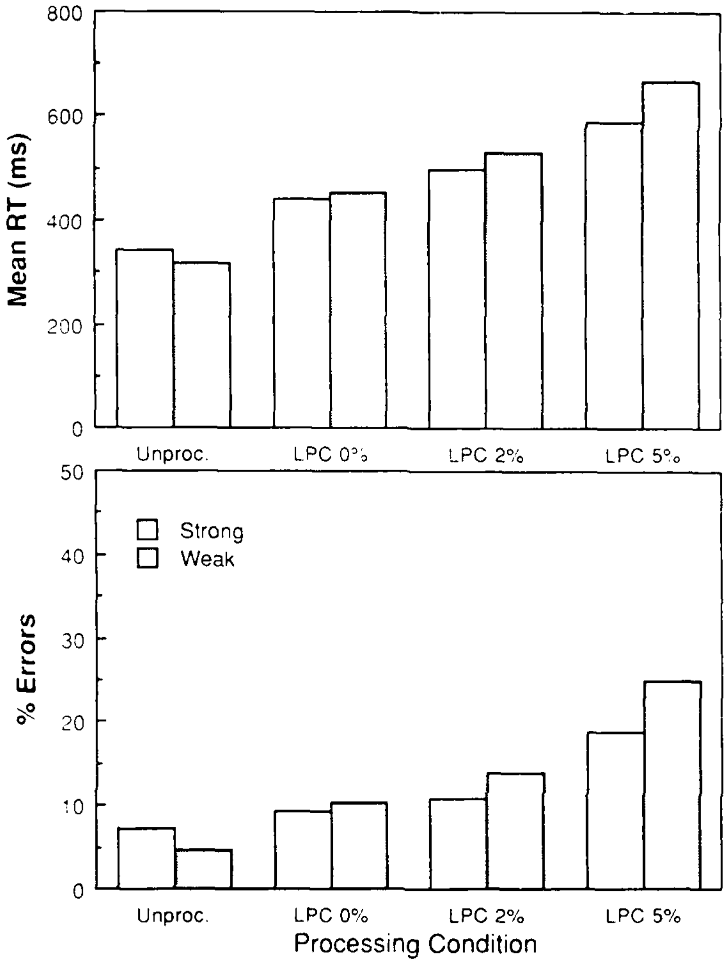


Fig. 1 — Performance as a function of subject-predicate relatedness (strong vs weak) and speech processing condition. Mean reaction time (RTs) are shown in the upper panel, and mean percentages of errors are shown in the lower panel.

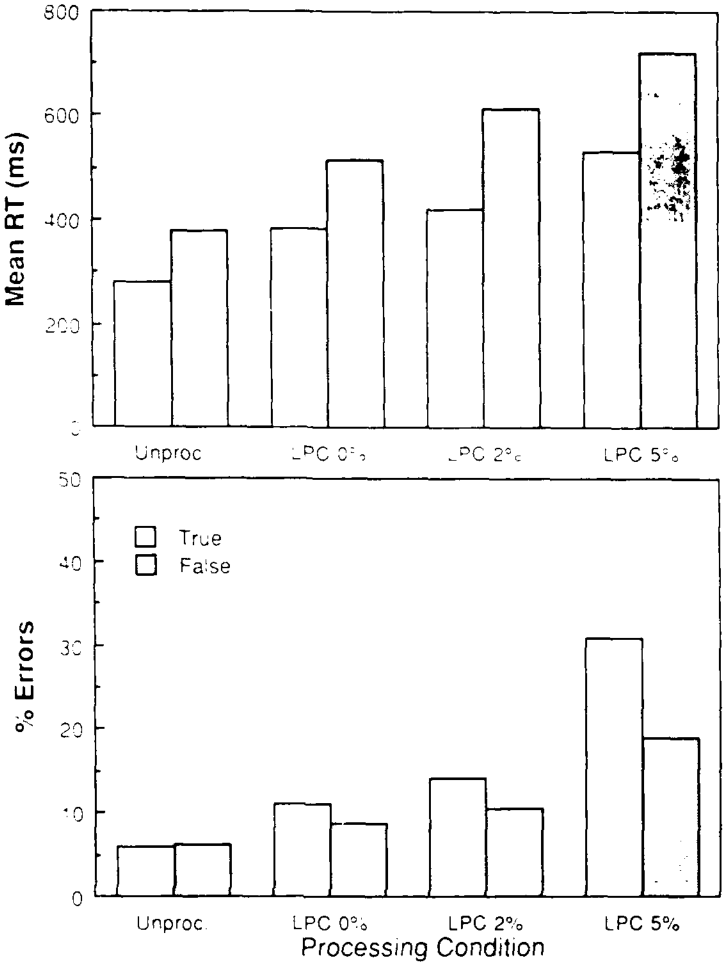


Fig. 2 – Performance as a function of truth value of the sentence and speech processing condition. Mean RTs are shown in the upper panel, and mean percentages of errors are shown in the lower panel.

The interactions involving subject-predicate relatedness and truth value were of particular interest. Although it was predicted that responses to true sentences would be faster and more accurate for strongly rather than weakly related sentences, a different set of predictions had been made for false sentences. Strongly related false sentences express relationships that can be difficult to distinguish from true ones. As a result, additional time would be required at the decision stage to respond to strongly related *false* sentences, even though word recognition may be facilitated because of priming by the strongly related early part of the sentence. Furthermore, because strongly related false sentences express relationships that are harder to distinguish from similar true ones (some listeners may not know for certain whether or not a camel has horns), the error rates for these sentences should be higher than would be predicted on the basis of intelligibility difficulties alone. The fact that the error rate for strongly related false sentences was 11.8% when *unprocessed* speech was presented supports this proposition. Because the error rate for weakly related falses was only 0.5%, it can be assumed that unprocessed speech provides little in the way of intelligibility difficulties, and the difference must be attributed to errors made at the decision stage.

As predicted, reaction times were faster (321 vs 487 ms) and error rates were lower (9% vs 22%) for strongly than for weakly related *true* sentences, whereas reaction times were faster (499 vs 615 ms) and error rates were lower (5% vs 14.2%) for weakly rather than for strongly related *false* sentences. The relatedness by truth value interactions were significant for reaction time, $F(1, 47) = 278.56$, $p < .001$, $MSe = 42,965$, as well as for errors, $F(1, 47) = 163.02$, $p < .001$, $MSe = 288.59$. The three way interaction of truth value, subject-predicate relatedness, and processor was not significant for the reaction times, $F < 1$, $MSe = 28,325$, but it was for the errors, $F(2.74, 128.79) = 3.59$, $p < .02$, $MSe = 141.11$ (Fig. 3). For true sentences, the effect of relatedness increased as the degradation of the speech increased, thus reflecting the increased value of contextual information as the speech became progressively degraded. In contrast, the advantage of weakly over strongly related false sentences *decreased* with increases in speech degradation, a result that also reflects the increased value of context as the speech degradation increased.

From the first replication to the second, mean reaction times decreased from 525 to 436 ms, $F(1, 47) = 53.39$, $p < .01$, $MSe = 59,944$, and replication did not interact with processing conditions for reaction time, $F(2.36, 111.05) = 1.11$, $p > .10$, $MSe = 45,820$. The mean error rate decreased from 13.7% to 11.4% from the first to the second replication, $F(1, 47) = 9.85$, $p < .01$, $MSe = 2$[illegible]$.09$, and the effect of processing condition on errors was smaller for the second than for the first replication, $F(2.66, 125.04) = 4.52$, $p < .01$, $MSe = 168.18$ (Fig. 4).

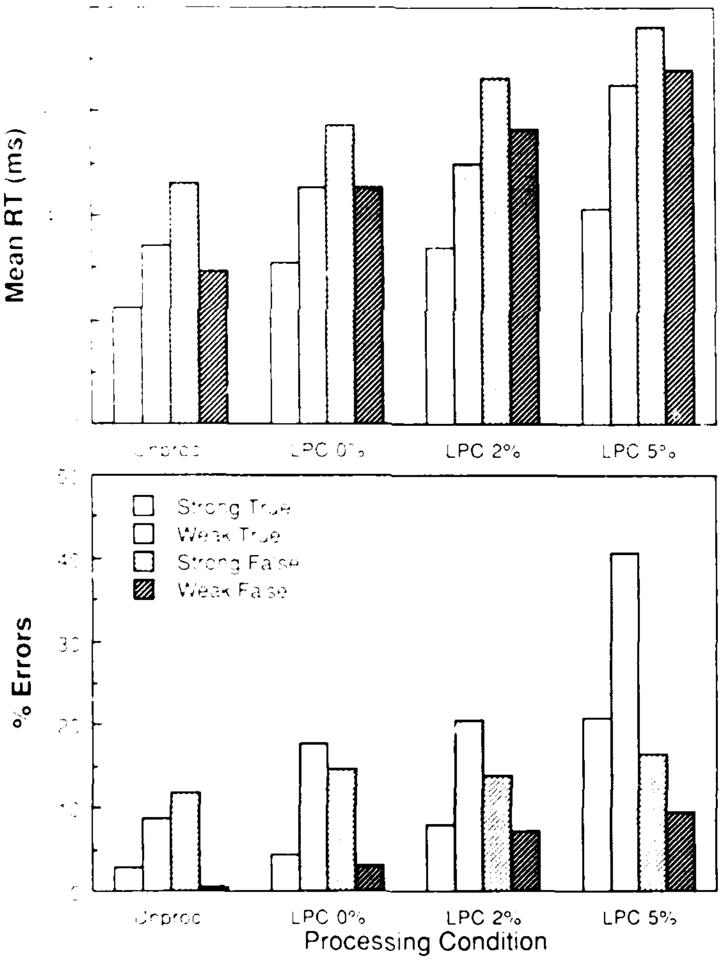


Fig. 3 — Performance as a function of subject-predicate relatedness, the truth value of the sentence, and speech processing condition. Mean RTs are shown in the upper panel, and mean percentages of errors are shown in the lower panel.

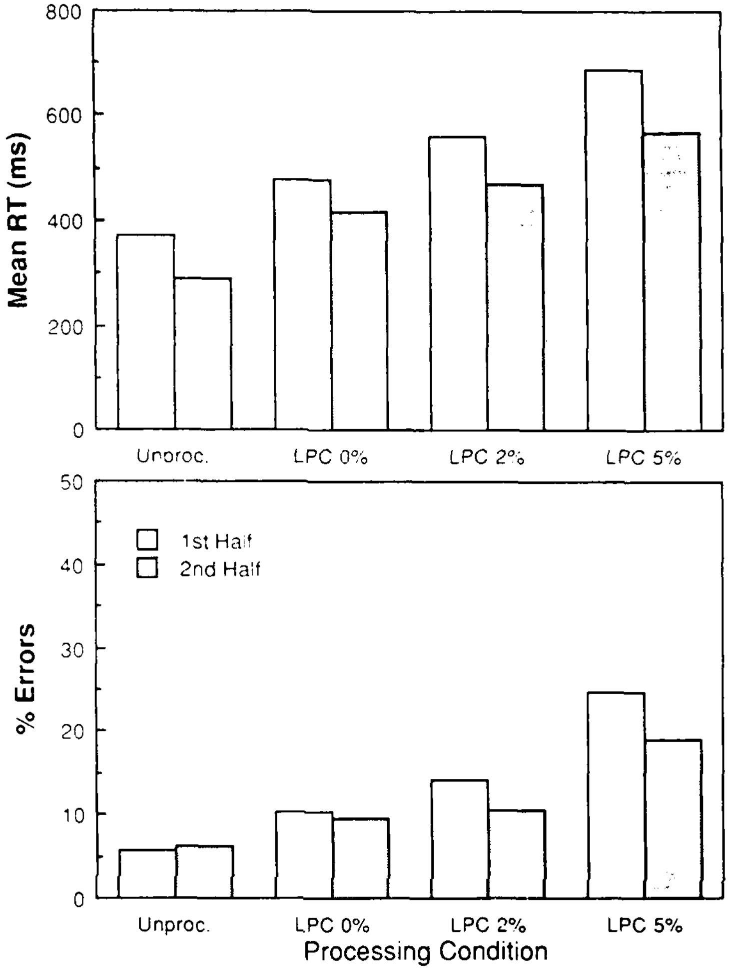


Fig. 4 — Performance as a function of replication (1st half vs 2nd half of experiment) and speech processing condition. Mean RTs are shown in the upper panel, and mean percentages of errors are shown in the lower panel.

The three way interaction of processor, subject-predicate relatedness, and replication was significant for the reaction times, $F(2.75, 129.04) = 4.53$, $p < .01$, $MSe = 22{,}019$, and for errors, $F(3, 141) = 5.27$, $p < .01$, $MSe = 124.79$ (Fig. 5). During the first replication, the effect of processor did not differ for weakly related and for strongly related sentences. In contrast, during the second replication, the effect of processor was greater for the weakly related sentences. Apparently, context was used more effectively to overcome speech degradations after listeners had become relatively practiced on the task.

Relationship to Previous Results

A comparison of the present results to previously obtained DRT and ICAO spelling alphabet scores is shown in Table 1. Although the low number of data points precludes the drawing of strong inferences about the functional relationships between the different measures of speech quality, it nevertheless appears that within the range of tested values, an increase of one point in the DRT results in a decrease in reaction time on the order of 10 to 20 ms in sentence verification, depending on factors such as the level of context.

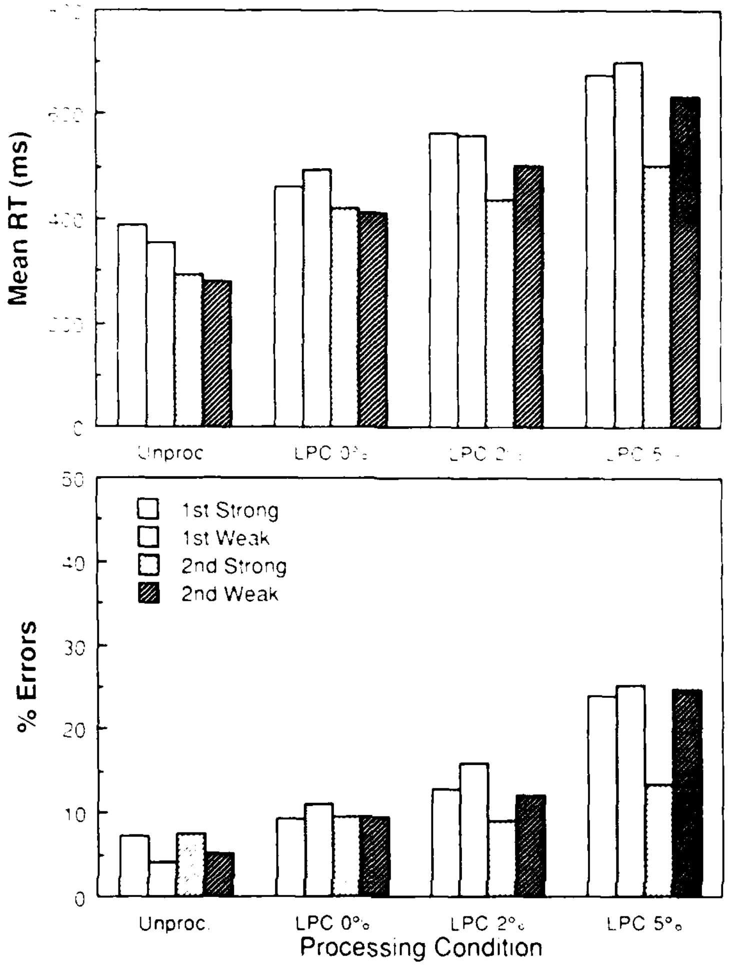


Fig. 5 — Performance as a function of subject-predicate relatedness, replication, and speech processing condition. Mean RTs are shown in the upper panel, and mean percentages of errors are shown in the lower panel.

Table 1 — Comparison of the Results of the Present Experiment with Previously Obtained DRT Scores and Percent Correct Responses on the ICAO Alphabet

Processing Condition	DRT Score	ICAO Score	% Correct Strong[a]	% Correct Weak[a]	Mean RT Strong[a]	Mean RT Weak[a]
Unprocessed	97.6	99.0	92.7	95.4	341	318
LPC 0% errors	86.0	98.0	90.6	89.6	442	455
LPC 2% errors	81.9	96.2	89.1	86.1	501	532
LPC 5% errors	75.4	90.3	81.2	75.0	588	667

[a] *Strong* and *Weak* refer to average scores for the strongly related and weakly related sentences.

DISCUSSION

The present experiment used a sentence verification task to test the comprehension of digitally processed sentences using the DoD standard LPC 2.4 kbps algorithm with and without random bit errors. The processed speech conditions tested here had been previously evaluated by using the DRT test and the ICAO spelling alphabet [4, 5]. The current approach, which required that the responses of listeners be based on comprehension of the content of each message, was motivated by the desire to obtain additional information about the effects of LPC processing and bit errors on speech effectiveness in the "real world." The results with the sentence verification task were systematic and interpretable within the framework that we outlined in the *introduction*.

Not surprisingly, reaction times and errors increased with increases in speech degradation. This was to be expected given previously obtained DRT scores and Pisoni and Dedina's finding [3] that LPC speech with no bit errors led to more errors and longer reaction times than wideband speech in a sentence verification task. The increased errors with LPC speech suggest that relatively inexperienced users may have some trouble when using LPC systems for ordinary, unconstrained conversational speech, although the improvement from the first half to the second half of the present experiment indicates that this difficulty could be reduced by practice. The large numbers of errors for LPC at the 2% and 5% bit error rates suggest that an open-ended vocabulary can be very difficult to understand under conditions of high bit errors. However, previously reported results using the ICAO spelling alphabet suggest that there would be substantially fewer comprehension errors when military or other constrained vocabularies are used.

Even in situations where the comprehension errors for a particular processor condition are at an acceptable level, the additional processing time required to understand the speech should also be taken into account in determining its acceptability. Our results suggest that the additional time required to comprehend a simple sentence when using LPC with 5% bit errors is on the order of 250 to 350 ms over that for an unprocessed sentence. This represents sufficient time for a typical adult to scan about seven digits in short-term memory [16] or to access four or five labels in long-term memory [17]. Therefore, it is probable that the additional time required to comprehend LPC speech with 5% bit errors would detract from other ongoing cognitive activities, a situation that might prove unacceptable if the listener has to respond quickly or engage in simultaneous tasks. Indeed, for some situations our estimates of the extra time required to comprehend LPC processed speech may be low. Pisoni and Dedina [3] found that reaction times using LPC at 2.4 kbps with no bit errors were more than 1 s longer than for 16 kbps wideband speech. Pisoni and Dedina's higher values for the additional time needed to comprehend LPC sentences may be the result of the minimal amount of practice they gave their listeners with LPC speech (i.e., exposure to only four sentences). Alternatively, it is possible that their speakers' voices were less suited to LPC processing than ours or that some other factor was responsible for the different results.

With increasingly degraded speech, reaction times were faster and more accurate for true sentences that expressed strong subject-predicate relationships than for weakly related true sentences, and the effect was greatest for the most severely degraded condition, LPC with 5% bit errors. It is reasonable to expect that the importance of context would increase with increasing degradation of the speech signal. With a high-quality speech signal, it should be possible to understand all of the words of a well-articulated sentence in the absence of contextual information. This contention is supported by the fact that in the unprocessed condition there were almost no errors for the weakly related false sentences, where there is little context to aid the comprehension stage. With unprocessed speech, perception could be based on data driven (or bottom-up) processes since, in this case, the speech data are sufficient to define the stimuli unambiguously. In contrast, with degraded speech the acoustic sensory information might not be sufficient for accurate perception of the words. Accordingly, conceptually driven (or top-down) processes would be required to *fill in* missing stimulus data. Because the critical words in the strongly related sentences are likely to prime one another, missing stimulus information would be compensated by knowledge based on context. The context of the weakly related sentences would be unlikely to prime or otherwise aid the recognition of words that might not be identifiable solely on the basis of the degraded stimulus information, and as a result, the performance would suffer.

In contrast to our results, Pisoni, Manous, and Dedina [18] in an experiment that tested the effect of sentence predictability on the perception of synthesized speech, found no interaction of predictability and speech type. The two studies are not directly comparable, however. For example they used high-quality synthetic speech as opposed to processed natural speech, and the differences in intelligibility among the various types of tested speech were greater in our study. Furthermore, they manipulated sentence predictability, whereas we manipulated subject-predicate relatedness. Whether either of these variables, or some other difference between experimental stimuli or procedures, was responsible for the different patterns of results is a question that can only be answered by further research.

It is well known that narrowband digital speech becomes easier to understand after practice, but we did not know how the improvement in performance (due to practice) would interact with the ability to use contextual information. As expected, faster reaction times and fewer errors were found in the second half of the experiment than in the first. The advantage of strongly over weakly related sentences in the degraded speech conditions was also greater in the second half than in the first half. It seems likely that while the LPC speech was still relatively novel, listeners needed to devote most of their attention to learning how to listen, and that this limited the attentional resources available to make use of contextual information. As the listeners became more familiar with the degraded speech, the mental processing of the speech may have become more automatic and attentional resources could then be freed to use the contextual information for top-down processing. This suggests that even though communicators experienced with LPC systems may perform very well in contexts in which they know what to expect; a novel or unexpected message could lead to errors and/or longer reaction times, especially in situations where the speech is further degraded by bit errors.

ACKNOWLEDGMENTS

The authors thank Corinne Meijer for her assistance in conducting the listener tests, Larry Fransen for his help with the digital processing, and Don Kallgren for lending us his speaking voice.

REFERENCES

[1] L.S. Larkey and M. Danly, "Fundamental Frequency and Sentence Comprehension," MIT Group Speech Working Papers, Vol. II, 1983.

[2] L.M. Manous, D.B. Pisoni, M.J. Dedina, and H.C. Nusbaum, "Comprehension of Natural and Synthetic Speech Using a Sentence Verification Task." Research on Speech Perception Progress Report 11, Indiana University, Bloomington, IN, 1985.

[3] D.B. Pisoni and M.J. Dedina, "Comprehension of Digitally Encoded Natural Speech Using a Sentence Verification Task (SVT): A First Report." Research on Speech Perception Progress Report 12, Indiana University, Bloomington, IN, 1986.

[4] A. Schmidt-Nielsen, "Intelligibility of ICAO Spelling Alphabet Words and Digits Using Severely Degraded Speech Communication Systems," NRL Report 9035, 1987.

[5] G.F. Sandy, *Digital Voice Processor Consortium Final Report MTR-84W00053*, Mitre Corp., McLean, VA, 1984.

[6] W.D. Marslen-Wilson and A. Welsh, "Processing Interactions and Lexical Access During Word Recognition in Continuous Speech," *Cognitive Psychology* **10**, 29-63 (1978).

[7] A. Salasoo and D.B. Pisoni, "Interaction of Knowledge Sources in Spoken Word Identification," *Memory and Language* **24**, 210-231 (1985).

[8] A.M. Collins, and M.R. Quillian, "Retrieval Time from Semantic Memory," *J. Verbal Learning and Verbal Behavior* **8**, 240-247 (1969).

[9] A.M. Collins, and M.R. Quillian, "Experiments on Semantic Memory and Language Comprehension," In *Cognition and Learning in Memory*, L.W. Gregg, ed., Wiley, New York, (1972) pp. 117-137.

[10] A.L. Glass, K.J. Holyoak, and C. O'Dell, "Production Frequency and the Verification of Quantified Statements," *J. Verbal Learning and Verbal Behavior* **13**, 237-254 (1974).

[11] E.H. Rosch, "Cognitive Representations of Semantic Categories," *J. Experimental Psychology: General* **104**, 192-233 (1975).

[12] W.F. Battig and W.E. Montague, "Category Norms for Verbal Items in 56 Categories: A Replication and Extension of the Connecticut Category Norms," *J. Experimental Psychology Monograph* **80**, (3, Pt. 2) (1969).

[13] R.F. Lorch, "Effects of Relation Strength and Semantic Overlap on Retrieval and Comparison Processes During Sentence Verification," *J. Verbal Learning and Verbal Behavior* **20**, 593-610 (1981).

[14] M. McCloskey and S. Glucksberg, "Decision Processes in Verifying Category Membership Statements: Implications for Models of Semantic Memory," *Cognitive Psychology* **11**, 1-37 (1979).

[15] H. Huynh and L.S. Feldt, "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Splitplot Designs," *J. Educational Statistics* **1**, 69-82, (1976).

[16] S. Sternberg, "High-speed scanning in human memory," *Science* **153**, 652-654 (1966).

[17] S.K. Card, T.P. Moran, and A. Newell, *The Psychology of Human-Computer Interaction*, Erlbaum, (Hillsdale, NJ 1983).

[18] D.B. Pisoni, L.M. Manous, and M.J. Dedina, "Comprehension of Natural and Synthetic Speech Using a Sentence Verification Task: II. Effects of Predictability on the Verification of Sentences Controlled for Intelligibility." Research on Speech Perception Progress Report 12, Indiana University, Bloomington, IN, 1986.