Published in final edited form as:

J Verbal Learning Verbal Behav. 1982 April ; 21(2): 186–195. doi:10.1016/S0022-5371(82)90551-5.

Some Effects of Perceptual Load on Spoken Text Comprehension

Hans Brunner and David B. Pisoni

Indiana University

Abstract

Investigations of spoken language comprehension often employ subsidiary task paradigms, such as word- and phoneme-monitoring procedures. In the present experiment we investigated what kinds of effects such secondary tasks might have on the course of simultaneous comprehension processing. Subjects listened to texts either without any secondary task demands (i.e., unconstrained comprehension), under word-monitoring instructions, or under phoneme-monitoring instructions. Comprehension was evaluated immediately after the presentation of each text with verification statements probing both the representation of surface structure and, also, the representation of high-level propositions, low-level propositions, and macropropositions from the text's meaning structure. Relative to unconstrained comprehension, we found an overall facilitation of comprehension due to word monitoring, and no effects as a result of phoneme monitoring. Moreover, the pattern of results across propositional levels was consistent with predictions derived from W. Kintsch and T. A. van Dijk's (*Psychological Review*, 1978, 85, 363–394) theory of text comprehension. This facilitation of simultaneous comprehension processing is inconsistent with the notion of a "fixed-capacity decision mechanism," to which differences in monitoring effects are typically attributed. We argue that subsidiary task paradigms can be used for investigations of spoken text comprehension, but only when accompanied with probes for the well-established characteristics of text representation.

Investigations of spoken language comprehension operate under unique paradigmatic constraints. The inherent distribution of spoken discourse across time eliminates any possibility of well-defined presentation intervals. Hence, it is impossible to obtain "on-line" measures analogous to the reading time and eye-movement data typically used in studies of reading (e.g., Hogaboam & Perfetti, 1975; Cirilo & Foss, 1980; Carpenter & Daneman, 1981).

Experiments on fluent speech comprehension have traditionally approached this problem by using what Levelt (1978) has recently referred to as "subsidiary task paradigms"; these are procedures that require subjects to focus attention on some aspect of a text's speech sound structure while simultaneously processing its semantic content. Inferences concerning the nature of on-line perception and comprehension processes are then drawn from the secondary task response latencies across different experimental conditions. Click detection (e.g., Abrams & Bever, 1969), shadowing (e.g., Rosenberg & Jarvella, 1970; Marslen-Wilson & Welsh, 1978), phoneme-monitoring (e.g., Bergfeld-Mills, 1980) and word-monitoring (e.g., Foss & Swinney, 1973) tasks are among the most frequently used variations of this approach.

By and large, these types of experimental paradigms have provided much useful information about the phonemic, lexical, and syntactic levels of language processing. However, their generality to the broader issues of spoken language comprehension has been limited by the implicit, and potentially false, assumption that their inherent secondary task characteristics have no effect on subjects' abilities to maintain otherwise normal comprehension

processing. Whether this is indeed the case is an almost completely untested assertion. Experiments using these monitoring techniques operate on the hypothesis of a fixed-capacity decision mechanism, the limitations of which are said to produce longer secondary detection latencies in response to increased text complexity. Given the way in which these experiments are typically conducted, however, there is also a strong, and often hidden assumption that speed—accuracy trade-offs are operating at only the “lower” more structural levels of language processing. Typically, however, this assumption is empirically unsupported. Experiments with adequate controls for enforcing or assessing criterial levels of comprehension are almost nonexistent, and often it is only assumed that the subject has processed the semantic content, without even explicitly instructing him or her to do so. Thus, in evaluating this line of research, there is no reason to believe that capacity limitations influence only the secondary detection latencies. Indeed, there is every reason to believe that they could well have significant effects on the primary comprehension process as well.

To date, there has been only one investigation of this issue, a recent study from our laboratory by Blank, Pisoni, and McClaskey (1981). Blank et al. had subjects listen to texts either without any secondary task demands (i.e., “normal,” unconstrained comprehension) or under conditions of word and phoneme monitoring. After listening to each text, however, the subjects were also required to answer a small set of multiple-choice questions concerning its content. The primary result of this study was a null finding; responses to these multiple-choice questions were just as accurate after comprehension in the two monitoring conditions as after normal, unconstrained comprehension. With these findings, the authors concluded that the secondary task demands of word- and phoneme-monitoring procedures appear to have no effects on simultaneous text processing.

The purpose of the present study was to expand on this initial finding and further evaluate the nature of spoken language comprehension under conditions of varying perceptual load. Blank et al. used a within-subjects design to manipulate their three monitoring conditions. Consequently, it is entirely possible that their null result could have been due to subjects adopting a single, low-level comprehension strategy rather than shifting strategies from one text to the next in response to fluctuating task demands. The conclusions of Blank et al. were also limited by imprecision in their measure of comprehension. The multiple-choice questions were presented in paper-and-pencil form, thereby eliminating any possibilities of more precise assessment through response latency measures. More seriously, however, there was also no systematic relationship between the specific test questions and the propositional structures of their experimental texts. The questions were simply chosen from previously published adult reading comprehension tests. This prevented the authors from going beyond the most fundamental question and ascertaining, in much greater detail, the *kinds* of effects that these perceptual tasks might have on spoken language comprehension.

In the present experiment we tried to redress these limitations by: (1) manipulating different comprehension conditions between subjects, (2) presenting comprehension probes over computer-driven CRTs, so that response latencies could be measured, and (3) systematizing the relationships between comprehension questions and the underlying propositional structures of the experimental texts. Four different kinds of questions, all of them requiring “yes/no” or “true/false” responses, were used to assess the different components of each text’s representation in memory: The representation of surface structure was assessed with *Remember* questions, asking subjects directly whether a randomly chosen word had occurred in the text or not. The distractors in this condition could be either words that rhymed or words that were synonymous with their appointed target words. Thus, in addition to gauging the overall strength of surface structure representation, an analysis of the relative false alarm rates to the semantic vs acoustically matched items should also tell us something

about the effects of comprehension conditions on the representation of lexical and surface information in the text.

Comprehension of semantic content was evaluated with three different kinds of true/false verification statements: The representation of each text's propositional microstructure was evaluated with verification statements focusing on either *high-level propositions* or *low-level propositions* (cf. Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975), explicitly present in the spoken text. Propositional macro structures were evaluated with *inference* statements, conveying information which could not be verified on the basis of simple fact retrieval alone, but which, instead, required some synthesis from the literally conveyed information. Presumably, different comprehension conditions should facilitate performance with such inferences only to the extent that they support the development of appropriate underlying macro-propositions (cf. Kintsch & van Dijk, 1978).

Method

Materials

The texts in this experiment were the same as those used in the earlier study of Blank et al. (1981). There were 12 experimental texts, each recorded by a female speaker onto one track of an audio tape with a professional-quality microphone (Unidyne HI, Model #545) and tape recorder (Ampex AG-500) in a sound-attenuated IAC booth. The 12 experimental texts were preceded by three practice stories. Each text was introduced with the spoken word, "Ready." Monitoring instructions, where necessary, were visually displayed on a CRT screen. A Latin-square design was used to produce three experimental tapes, each beginning with the same three practice stories and differing from the other two only in the serial order of its constituent experimental texts.

Comprehension was evaluated with true/false and "Remember" questions. Examples of these are shown, with one of the experimental texts, in Table 1. As can be seen, the presentation of each text was followed by four questions: The first kind of question, the *Remember* questions, always consisted of a standard sentence frame ("Did the word 'XXXXX' occur in this story?") with either a target (e.g., "look"), a rhyming distractor (e.g., "book"), or a synonymous distractor (e.g., "gaze") in the test position. Comprehension of propositional content was evaluated with verification statements pertaining either to *High-Level Propositions*, *Low-Level Propositions*, or emergent *Inferences* from the text. The high- and low-level probes tested for information explicitly conveyed in the text. Inference questions bore no direct relationship to explicit propositional information and could be answered only by realizing emergent relationships from the text's explicit propositional structure.

Procedure and Apparatus

Experimental sessions were conducted in groups of one to five subjects. Each subject was seated in an experimental cubicle equipped with a GBC Standard CRT display monitor (Model MV-10A), a seven-button response box, and a pair of TDH-39 headphones connected to an Ampex AG-500 tape recorder. Each booth was interfaced to a PDP-11/05 computer, which presented instructions and test questions on the CRT and recorded all responses and latencies from the response boxes.

Instructions were read to the subjects by the experimenter. Throughout this process, all references to the critical points of each trial were illustrated with display sequences of what would actually be seen during the experiment. Subjects were told that they would hear a number of short stories about various topics and that they were to listen to these for content and try to understand them. After being shown examples of the different kinds of questions

and trial prompts, subjects were then given a single practice trial without any monitoring task. Subjects in the normal comprehension condition were then given two more practice trials of the same form before starting on the actual experiment.

Subjects in the two monitoring conditions, however, were told that they would also have to listen for the presence of a pre-specified, word-initial target phoneme or target word during the presentation of each story. Monitoring targets would be visually displayed to them prior to the start of each story. During the story's presentation they were to keep their index finger resting lightly on a "READY" button (mounted in the center of each response box) and press it as quickly as possible whenever the pre-specified target was detected. Subjects in the monitoring conditions were then given two more practice trials (using the same stories heard by the normal comprehension group) with monitoring instructions appropriate to their particular condition.

Subjects listened to 12 experimental texts, as in Blank et al. (1981). The presentation of each story was visually prompted ("ATTENTION! New Story Coming Up. Please press READY to begin.") and delayed until all subjects had pressed their READY buttons. In the two monitoring conditions, new story prompts were augmented with a display of the trial's target word or phoneme (e.g., "Listen for the sound /d/, as in "dancing"). After hearing each text through headphones, the subjects were shown a display sequence of four randomly ordered questions. Each question was subject-initiated, and the response to it was followed by (1) a 7-point confidence rating on the answer just provided and (2) feedback (via lights mounted in each response box) indicating the correct answer for that question. The last question was then followed by a display signaling the start of the next trial.

Subjects and Design

The three conditions of comprehension (i.e., unrestricted listening, word monitoring, and phoneme monitoring) were manipulated as a between-subjects variable. The four kinds of questions were manipulated within subjects. For purposes of counterbalancing, two control variables (manipulated between subjects) were also included. These were tapes (three levels) and response side (i.e., whether "yes/true" was indicated by pressing a button to the right or left of center).

The data from 90 subjects, 30 in each listening condition, were required for this design. Subjects were all Indiana University undergraduates whose participation in the experiment served as partial fulfillment of a course requirement. All subjects were native speakers of English and reported no history of hearing or speech disorders prior to testing.

Results

Two a priori criteria were used to identify and eliminate unacceptable monitoring and question answering performances: To be included in the final analyses, subjects had to: (1) detect at least 60% of their monitoring targets and (2) perform at greater than chance level (50%) in their responses to the questions. Twenty-four percent of the total number of subjects run for this experiment ($N = 120$) failed to meet one of these two criteria.

This rejection rate, is shown, broken down across conditions, in Table 2. Examination of this table shows that the largest subject attrition rate occurred in the phoneme-monitoring condition. The proportion of subjects discarded for inadequate question answering after normal comprehension is also noticeable, being 75% larger than the corresponding rates for the two monitoring conditions. With respect to the elimination of "bad" subjects then, it would appear that the question-answering and detection criteria were sufficiently redundant for question answering by itself to produce almost as much subject selectivity as the two

criteria put together. The only other preparation performed on the data was elimination of response latencies standing more than three standard deviations away from their respective comprehension condition cell means.

Monitoring and question-answering data were examined with analyses of variance, alternately treating subjects and texts as the random variable (cf. Clark, 1973). Throughout this section, F_s will represent subjects-random F ratios and F_t text random F ratios. However, where significant, it is $\min F'$ which will be reported. Unless otherwise stated, all p 's reported in this section are less than .01.

Monitoring Data

Mean latencies and detection rates from the two monitoring conditions were computed, separately, for each subject and text. These means were based on reaction times ranging from 100 to 1200 milliseconds in length. Responses with latencies outside these limits were presumed to reflect either anticipation or inattention and were therefore excluded from further analysis.

The overall word- and phoneme-monitoring condition means are shown in Table 3. As can be seen, words were detected more quickly and accurately than phonemes. The 110-millisecond difference ($\min F'(1,61) = 26.1$) in monitoring latencies and the 9% difference ($\min F(1,61) = 16.62$) in monitoring detection rates were both statistically reliable.

This difference in monitoring performance is entirely consistent with findings reported previously in the literature (e.g., Savin & Bever, 1970; Foss & Swinney, 1973; Blank et al., 1981). Our overall detection levels, however, were somewhat lower than those typically reported in the literature (e.g., Foss & Swinney, 1973; Bergeld-Mills, 1980). We can account for this with two observations: First, most previous monitoring studies have used lists of individually presented sentences rather than coherent, fluent texts. The extraction of a text's emergent meaning adds considerably to on-line resource demands and could certainly act as a legitimate cause for lower secondary performance. Second, except for the Blank et al. work, no previous studies have emphasized comprehension processing with immediate text probes. The presence of such questions in the current experiment undoubtedly forced subjects to maintain certain minimal levels of comprehension performance. Lower detection rates seem to have resulted from this; an observation which underscores our concern over the assumptions which have been made about simultaneous comprehension processing in these kinds of experiments in the past.

Since our primary aim was to assess the relative effects of these two secondary tasks on the course of ongoing comprehension, it was important to ensure that both tasks had been performed at the same level of proficiency. The current difference in detection rates indicates that, despite our best efforts with a priori rejection criteria, we were unsuccessful in accomplishing this. Consequently, it was necessary to analyze the question-answering data in two forms; once in raw form and once again after removal of the covariance (cf. Kerlinger & Pedhazur, 1973) due to these differences in monitoring detection rates.

Representation of Surface Structure

The effects of monitoring tasks on the processing of surface structure were evaluated by examining subjects' performances on the Remember questions. Overall, differences in monitoring conditions had no significant effects on either the proportion of correct Remember target identifications (mean % correct = 0.83) or on subjects' rated confidence in their identification responses (mean rating = 5.34). However, there was a reliable main effect of monitoring conditions on the correct Remember target identification latencies ($F_s(2,87) = 3.83$, $MS_{\text{err}} = 874,298$, $p < .05$; $F_t(2,22) = 7.78$, $MS_{\text{err}} = 227,381$). Monitoring

performances covaried significantly with this measure ($F_s(1,71) = 4.3$, $MS_{\text{err}} = 56,802$, $p < .04$; $F_t(1,23) = 4.86$, $MS_{\text{err}} = 187,228$, $p < .038$). Nonetheless, the effect remained marginally significant after removal of this covariance ($F_s(2,86) = 2.96$, $MS_{\text{err}} = .01$, $p < .06$; $F_t(2,32) = 3.7$, $MS_{\text{err}} = .021$, $p .05$). The adjusted cell means are, normal comprehension: 3503 milliseconds; word monitoring: 3139 milliseconds; phoneme monitoring: 3540 milliseconds. Thus, subjects were fastest at target detection after word monitoring and slower, by roughly equal amounts, after unconstrained comprehension and comprehension with phoneme monitoring.

Analysis of the false alarm rates for the two forms of distractor items (rhyming vs synonymous) produced a reliable main effect across distractor type ($\min F'(1,15) = 12.56$), but none across monitoring conditions. This, too, was replicated after removal of the covariance due to monitoring performance ($\min F'(1,112) = 22.15$). Rhyming distractors ($p(\text{correct rejection}) = .702$) were rejected almost twice as accurately as synonymous distractors ($p(\text{correct rejection}) = .368$). However, there were no reliable effects on distractor confidence ratings (mean = 3.62) or response latencies (mean = 3925 msec).

Selective attention to the speech sound structure of spoken language input might have produced qualitative differences in the representation of surface structural information. This, however, was not observed in our analyses since there were many more false alarms to distractors with semantically matching information. Thus, regardless of perceptual demands, the encoding of surface structural information occurred rapidly and appeared uniformly semantic in nature (cf. Jarvella, 1971).

Representation of Meaning

The representation of prepositional microstructure was evaluated by examining subjects' performances on the high- and low-level proposition verification statements. Probabilities of correct response to these questions were unaffected by monitoring conditions and propositional level ($p(\text{correct response}) = .877$). Levels of confidence, however, were reliably influenced. Analysis of the raw data produced a significant propositional levels effect ($F_s(1,187) = 23.7$, $MS_{\text{err}} = .254$; $F_t(1,11) = 4.86$, $MS_{\text{err}} = .289$, $p < .05$), which was also replicated after equating for differences in monitoring detection performance ($\min F'(1,336) = 3.96$, $p < .05$). The adjusted cell means for rating responses were high-level propositions: 6.06; low-level propositions: 5.69. Thus, despite equivalent performance levels, subjects had more difficulty retrieving and verifying ideas conveyed at the low levels of a text's micropropositional structure than they did with ideas comprising the higher levels of its text base. Of course, this failed to interact in any way with the secondary perceptual load accompanying comprehension.

A levels effect was also obtained across high- and low-level proposition verification latencies. Again, this was evident in both raw ($F_s(1,87) = 22.2$, $MS_{\text{err}} = 530,809$; $F_t(1,33) = 4.74$, $MS_{\text{err}} = 498,479.9$, $p < .05$) and covariance-adjusted ($F_s(1,175) = 12.78$, $MS_{\text{err}} = .005$; $F_t(1,66) = 4.69$, $MS_{\text{err}} = .013$, $p < .05$) forms of the data. This time, however, there was a significant main effect of monitoring conditions ($F_s(2,175) = 3.87$, $MS_{\text{err}} = .005$, $p < .05$; $F_t(2,66) = 3.77$, $MS_{\text{err}} = .013$, $p < .05$), which was present only in the covariance-adjusted response latencies. The adjusted cell means are shown in Table 4. As can be seen in Table 4, high- and low-level propositions were verified more quickly after comprehension under word monitoring than after unconstrained comprehension or comprehension with phoneme monitoring. From the superscripts, however, it appears that the effect is somewhat larger for high-level propositions than low-level propositions. Although a significant interaction failed to obtain, covariance analyses performed separately on the two levels of proposition support this observation: A reliable monitoring conditions main effect was obtained for high-level propositions ($F_s(2,86) = 3.7$, $MS_{\text{err}} = .011$, $p < .05$; $F_t(2,32) = 3.74$, $MS_{\text{err}} = .027$, $p < .05$),

but not for the latencies of low-level propositions. Thus, one may tentatively conclude that the facilitation due to word monitoring is greater if the to-be-retrieved information is stored at higher levels of a text's micro-propositional structure than lower levels.

The development of propositional macrostructure was evaluated from performances on the inference questions. Confidence ratings (mean = 5.4) and the probabilities of a correct response (mean = .76) to these questions were unaffected by the three levels of secondary task. The response latencies, however, reflected a significant facilitation due to word monitoring in both raw ($F_s(2,87) = 3.07$, $MS_{\text{err}} = 1,892,965$, $p < .051$; $F_t(2,22) = 8.1$, $MS_{\text{err}} = 269,151$) and covariance-adjusted ($F_s(2,86) = 3.13$, $MS_{\text{err}} = .011$, $p < .05$; $F_t(2,32) = 7.4$, $MS_{\text{err}} = .02$) forms of the data. The adjusted cell means, shown in Table 4, reveal a trend quite similar to that shown for the high-level propositions across monitoring conditions. However, there were also some substantial differences: Inference questions took 1100 milliseconds (26%) longer to respond to than high- and low-level propositions ($\min F(1,40) = 20.38$). This was also the only condition to produce any kind of noticeable difference between performances after unconstrained comprehension and comprehension with phoneme monitoring. Although statistically unreliable, there was a 240-millisecond difference between the cell means of these two conditions, with phoneme monitoring producing longer response latencies to these questions.

Conclusions

At the onset of this study we had anticipated that subsidiary task demands, if they had any effects at all, would surely operate to decrease performance with simultaneous comprehension processing. Clearly, we were wrong in this prediction. On the whole, the secondary tasks of this experiment had relatively small effects on the course of ongoing comprehension. Contrary to what one might have expected, selective attention to the speech sound structure did not improve the encoding of phonemic representations of surface structure, nor did it have any effect, in any of the conditions, on the accuracy of subjects' responses. Indeed, we essentially replicated the earlier finding of Blank et al. (1981) of a null relation between monitoring tasks and the accuracy of question answering performance. At one level then, "on-line" processing capacities seem not to have been exceeded by these manipulations. Increases in perceptual demands due to secondary target monitoring did not produce commensurate decreases in the quality of simultaneous comprehension processes. Hence, on the basis of this evidence, the assumed restriction of speed-accuracy trade-offs to secondary levels of processing seems to have been supported.

This conclusion, however, requires the assumption that comprehension and secondary target monitoring processes were operating independently of each other—and thereby contradicts the consistent facilitation due to word monitoring which we found. The fact that word monitoring did facilitate comprehension indicates that monitoring and comprehension processes did *not* operate independently of one another and that, therefore, the assumption of processing independence would appear to be an invalid one.

There are two points in the target-monitoring paradigm at which facilitation of this kind might have occurred. The first is in the sensory/semantic identification of individual words (i.e., lexical access). The debate on whether words or phonemes are more fundamental to the perception of fluent speech has been carried on for some time now (e.g. Savin & Bever, 1970; Foss & Swinney, 1973; McNeill & Lindig, 1973). The current wisdom seems to favor "dual code" hypotheses (e.g., Foss, Harwood, & Blank, 1980; Dell & Newman, 1980), positing the identification of words on the basis of both lexical and phonetic information. If words were identified through serial constituent analyses of individual phonemes from the acoustic/phonetic input, then selective attention to the phonemic level of incoming speech

should have had a much larger effect on comprehension than selective attention for whole words. However, the opposite finding obtained. Thus, if our effects were due to some facilitation at the level of word identification, then they would indicate that the point of sensory/ semantic interaction contributes more to the phonetic analysis of whole words, than to constituent analysis of their phonemic components (Marslen-Wilson & Welch, 1978).

Selective attention to whole words might also have influenced the integration of text structure. This explanation is supported by the close match between our results and predictions which might be derived from Kintsch and van Dijk's (1978) model of text comprehension. According to this theory, the coherence of a text is established through a three-step cycle in which ordered lists of new propositions are integrated with the preceding text via argument overlap with a limited set of specially marked, or "reinstated," anchor propositions. This is said to occur at both the -micro- and macropropositional levels of meaning, with the former process operating on sentences and phrases, and the latter at the paragraph level. The model makes two critical assertions; namely, (1) that the (micro)propositionalization of new input occurs independently of the integration cycle and (2) that, once propositionalized, the strength of an idea's representation in memory should increase as a function of the number of times it has been reinstated into the cycle. Thus, since they are more central to the overall meaning of a text, superior memory for macropropositions and high level micro-propositions can be accounted for on the basis of the higher number of reinstatements which they have putatively undergone.

However, if this is the case, then it also follows that processes which facilitate or interfere with the reinstatement cycle should have disproportionately larger effects on the subsequent representation of higher-level propositions. This is precisely what obtained in the present results. From Table 4, it can be seen that the facilitory effects of word monitoring reduced the response latencies for low-level propositions by 173 milliseconds and those for high level proposition and inference questions, by 560 milliseconds. Thus, the effects of monitoring conditions were indeed larger for higher levels of the text base. According to the Kintsch and van Dijk (1978) model then, a text integration account can also be given for the present results: Selective attention to whole words facilitates comprehension by increasing the number of reinstatement cycles that central or high-level ideas are applied to, relative to ideas presented in the lower levels of the text base.

In conclusion, we would argue against the use of subsidiary task paradigms by themselves as a valid method for investigating on-line processes of spoken language comprehension. The finding of enhanced comprehension in response to secondary demands is inconsistent with the metaphor of a fixed-capacity decision mechanism operating on relatively homogeneous input. It is not so much the notion of limited capacity which our data contradicts but rather the assumption of a "relatively homogeneous input". Secondary task demands appear to interact strongly with the object of selective attention. In the present study we have seen that words and phonemes are not equally "foreign" to the development or encoding of propositional levels of representation. As a consequence, there are important differences in the kinds of effects that the two secondary tasks have on simultaneous comprehension processing. Subsidiary task paradigms provide only half the picture of what occurs during comprehension. Although useful by themselves for investigations of "on-line" perceptual processing, such techniques should be augmented with well-established measures of text representation when the object of study concerns spoken comprehension. Additional work is currently underway on these problems in our laboratory.

Acknowledgments

This research was supported, in part, by NIMH Research Grant MH-24027, NIH Research Grant NS-12179, and NIH Training Grant NS-07134 to Indiana University. Address reprint requests to Hans Brunner or David Pisoni, Department of Psychology, Indiana University, Bloomington, Ind. 47405.

References

- Abrams K, Bever TG. Syntactic structure modifies attention during speech perception and recognition. *Quarterly Journal of Experimental Psychology*. 1969; 21:280–290. [PubMed: 5347013]
- Bergfeld-Mills C. Effects of context on reaction time. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:75–83.
- Blank MA, Pisoni D, Mcclasky CL. Effects of target monitoring on understanding fluent speech. *Perception & Psychophysics*. 1981; 29:383–388. [PubMed: 7279562]
- Carpenter PA, Daneman M. Lexical retrieval and error recovery in reading: A model based on eye fixations. *Journal of Verbal Learning and Verbal Behavior*. 1981; 20:137–160.
- Cirilo RK, Foss DJ. Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:96–109.
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:335–339.
- Dell GS, Newman JE. Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:608–623.
- Foss DJ, Swinney DA. On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:246–257.
- Foss, DJ.; Harwood, DA.; Blank, MA. Perception and production of fluent speech. Hillsdale, NJ: Erlbaum; 1980. Deciphering decoding decisions: Data and devices.
- Hogaboam TW, Perfetti CA. Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:265–274.
- Jarvella RJ. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*. 1971; 10:409–416.
- Kerlinger, FN.; Pedhazur, EJ. Multiple Regression in Behavioral Research. New York: Holt: Rinehart & Winston; 1973.
- Kintsch W, Kozminsky E, Streby W, Mc-Koon G, Keenan JM. Comprehension and recall of a text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*. 1975; 14:196–214.
- Kintsch W, Van Dijk TA. Toward a model of text comprehension and production. *Psychological Review*. 1978; 85:363–394.
- Levelt, WJM. A survey of studies in sentence perception: 1970–1976. In: Levelt, WJM.; Flore-d'Arcais, GB., editors. *Studies in the perception of language*. New York: Wiley; 1978.
- Mcneill D, Lindig K. The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:419–430.
- Marslen-Wilson WD, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*. 1978; 10:29–63.
- Morton J, Long J. Effect of word transitional probability on phonemic identification. *Journal of Verbal Learning and Verbal Behavior*. 1976; 15:43–51.
- Rosenberg S, Jarvella RJ. Semantic integration and sentence perception. *Journal of Verbal Learning and Verbal Behavior*. 1970; 9:548–553.
- Savin HB, Bever TG. The non-perceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*. 1970; 9:295–302.

TABLE 1

Prototypical Text with Comprehension Probes

The distance to any nearby star is measured by observing how much it appears to change its position in an interval of six months against a background of more remote stars. This method can be illustrated if you look at one finger held at arm's length before your eyes. Close one eye and sight at the background. Then quickly close that eye and open the other, not moving the finger. Your finger appears to jump some distance before your eye. Yet you have not moved it. The amount that your finger jumps is large and can be measured in angles because the base line is the distance between your eyes. A star, because it is so far away, appears to move a very small amount against the background of stars in the distance. The observations are made six months apart in order to secure a very large baseline—the length of the earth's orbit, which is equal to 186,000,000 miles.

Remember question

“gaze” (S)

Did the word “look” (T) occur in this story?

“book” (R)

High-level proposition

The distance to a star is measured from its apparent motion.

Low-level proposition

In the demonstration, the baseline is the distance between your eyes.

Inference statement

If your finger were moved further away, it would jump a smaller distance.

Note: “T” indicates the target word, “S” synonym distractors, and “R,” rhyming distractors.

TABLE 2

Proportions of Subjects Discarded due to Inadequate Monitoring and Question-Answering Performance

	Normal comprehension	Word monitoring	Phoneme monitoring
Monitoring	—	.029	.086
Question- answering	.043	.029	.029

Note. Proportions have been computed against the total number of subjects run for the experiment ($N = 120$).

TABLE 3

Mean Latencies and Detection Rates for Word- and Phoneme-Monitoring Conditions

	Word monitoring	Phoneme monitoring
Latencies (msec)	534	646
Detection rates	.874	.782

TABLE 4

Covariance-Adjusted Mean Response Latencies (msec) for Inference, High-Level, and Low-Level Proposition Verification Statements

	High level	Low level	inference
Normal comprehension	4236b	4603 ^c	5743 ^e
Word monitoring	3670 ^a	4430 ^{bc}	5183 ^d
Phoneme monitoring	4201 ^b	4636 ^c	5987 ^e

Note. Cell means with the same superscript do not differ significantly ($p < .05$).