Published in final edited form as:
Percept Psychophys. 1981 April ; 29(4): 383–388.

Effects of target monitoring on understanding fluent speech

MICHELLE A. BLANK, **DAVID B. PISONI**, and **CYNTHIA L. McCLASKEY**
Indiana University, Bloomington, Indiana 47405

Abstract

Phoneme monitoring and word monitoring are two experimental tasks that have frequently been used to assess the processing of fluent speech. Each task is purported to provide an "on-line" measure of the comprehension process, and each requires listeners to pay conscious attention to some aspect or property of the sound structure of the speech signal. The present study is primarily a methodological one directed at the following question: Does the allocation of processing resources for conscious analysis of the sound structure of the speech signal affect ongoing comprehension processes or the ultimate level of understanding achieved for the content of the linguistic message? Our subjects listened to spoken stories. Then, to measure their comprehension, they answered multiple-choice questions about each story. During some stories, they were required to detect a specific phoneme; during other stories, they were required to detect a specific word; during still other stories, they were not required to monitor the utterance for any target. The monitoring results replicated earlier findings showing longer detection latencies for phoneme monitoring than for word monitoring. Somewhat surprisingly, the ancillary phoneme-and word-monitoring tasks did not adversely affect overall comprehension performance. This result undermines the specific criticism that on-line monitoring paradigms of this kind should not be used to study spoken language understanding because these tasks interfere with normal comprehension.

When faced with the task of understanding spoken language, listeners are rarely conscious of the sound structure of an utterance. The primary focus of the listeners' conscious awareness of the speech signal is directed toward understanding the content of the message and not toward analyzing its constituent elements (viz., the individual phonemes, syllables, or words). Despite this observation, subjects are nevertheless able to make reliable judgments about the detailed properties of the sound structure of an utterance while at the same time also devoting efforts toward comprehending the message. This has been demonstrated numerous times in experiments using paradigms such as phoneme monitoring, word monitoring, and mispronunciation detection (see, e.g., Cole & Jakimik, 1980; Foss & Blank, 1980; Marslen-Wilson& Tyler, 1980).

Phoneme monitoring, word monitoring, and mispronunciation detection are representative of a class of experimental techniques that have been used quite often to assess various components of fluent speech comprehension. Each of these tasks has been assumed to provide a measure of ongoing comprehension processes. All involve latency measures and all are assumed to index "momentary processing load" during fluent speech perception. Each task explicitly entails directing the subjects' attention to some property of the sound structure of the speech signal while at the same time requiring listeners to comprehend the utterance. (See Levelt, 1978, for an extensive review of studies using tasks of this kind.)

Reprints may be obtained from either Michelle Blank, who is now at Bell Laboratories, Whippany, New Jersey 07981, or David Pisoni at Indiana University.

Even though these tasks have been used extensively in the past, only recently has an interest developed in specifying the perceptual and cognitive processes involved in the tasks themselves (see, e.g., Blank, 1980; Foss & Blank, 1980; Rudnicky, 1980). Other than some speculation among theorists, there is still relatively little known about the effects of the task demands of target monitoring on comprehension performance. This lack of knowledge on our part is by no means trivial. After all, task demands in psychological experiments, particularly experiments involving linguistic materials, have been shown to affect the perceptual organization and encoding of the stimuli (e.g., Aaronson, 1976; Ammon, Ostrowski, & Alward, 1971; Carey, 1971). Thus, the validity of phoneme-monitoring, word-monitoring, and mispronunciation-detection tasks as measures of ongoing language comprehension would appear to be limited without much more detailed knowledge about how these tasks affect the normal processes of spoken language understanding. We would not want to base our theoretical accounts of spoken language comprehension on experimental paradigms that may disrupt the integrity of the very processes we wish to study.

In this paper, we are interested in the following specific issue: Does the conscious allocation of processing resources to different levels of the sound structure of the speech signal affect the way listeners process the utterance and the ultimate level of understanding achieved for the content of the message? The experiment reported here is a preliminary investigation that examines the potential interfering effects of phoneme monitoring and word monitoring on the normal comprehension process. Suppose we find that comprehension is impaired by performing the ancillary task of monitoring the speech signal for a phoneme or word throughout a passage of connected discourse. This result would then have to be taken into account when drawing inferences about language comprehension from monitoring data of this type. If we find deleterious effects of phoneme and word monitoring on comprehension in the present experiment, then we will have empirical justification to extend our criticisms to other popular on-line measures of fluent speech decoding, such as mispronunciation detection and speech shadowing.

By examining listeners' performances on various kinds of comprehension questions as a function of different monitoring conditions, we hope to learn something about the specific task demands of phoneme and word monitoring and how they may interact with ordinary comprehension processes. Of primary interest in the present study is the comparison of comprehension performance in the two monitoring conditions, on the one hand, with a nonmonitoring control condition, on the other hand.

The other important question that this study addresses is whether monitoring at the word level will have the same effects on comprehension as monitoring at the phoneme level. The answer to this question bears directly on the roles of lexical and phonemic representations in speech processing. Several theorists have argued that the computation of phonemic information is not a basic, or even necessary, process in the perception and comprehension of fluent speech (Klatt, 1979; Marslen-Wilson & Tyler, 1980; Morton & Long, 1976; Warren, 1976). Instead, these investigators claim that lexical, and not phonemic, representations play a primary role in understanding spoken language. If this view of the primacy of lexical interpretation in speech understanding is legitimate, then we might expect phoneme monitoring to interfere with ongoing comprehension processes more than would ivord monitoring. On the other hand, if phonemic as well as lexical representations are normally computed during fluent speech processing (see Foss & Blank, 1980; Blank, Note 1), then both word- and phoneme-monitoring tasks might be expected to have more or less comparable effects on comprehension processes. Both targeting conditions would, nevertheless, be expected to produce selective decrements in comprehension performance when compared with the nonmonitoring control condition. Both targeting tasks *explicitly*

require a listener to make an overt response about a specific property or attribute of the sound structure of the speech signal that is not typically brought to conscious awareness during the usual course of sentence processing and language comprehension activities. A monitoring task that requires explicit attention to sound attributes may promote and perhaps even require the use of special perceptual and cognitive strategies. This, in turn, may adversely affect comprehension in ways that are currently unknown.

METHOD

Design

Twelve narrative stories were chosen from several published adult reading or listening comprehension tests. (See Table 1 for the exact details of the passages.) In order for each of the stories to occur in the three experimental conditions (viz., nonmonitoring, word monitoring, and phoneme monitoring), three sets of tapes were constructed. Each set contained all 12 stories; 4 of the stories in the three sets came from each of the three conditions. The experiment was therefore a 3 (monitoring: none/word/ phoneme) by 3 (tape sets) factorial design, with the former variable within-subjects and the latter between-subjects. The comprehension test questions for each story were identical across the three tape sets.

Materials

A female speaker (M.A.B.) recorded all 12 stories on one track of an audiotape with a professional quality microphone and tape recorder in a sound-attenuated IAC booth. Each story on this master tape was preceded by the word “Ready” and three target specifications (viz., “Do not listen for any target”; “Listen for the target word_____”; “Listen for the target sound_____”). Initially, 4 (of the total 12) stories were assigned randomly to each of the monitoring conditions. Then, using a roll-over design, the three tape sets were made by cross-recording the master tape and editing the target specifications so that each story occurred in each monitoring condition across the tape sets.

Presentation of the stories for each tape set was blocked by monitoring condition. Order of presentation of the monitoring condition was counterbalanced for each tape set. Thus, a tape set consisted of three tapes that differed only in the order of condition presentation. A total of nine tapes were cross-recorded.

The phoneme targets in the phoneme-monitoring condition consisted of all and only the word-initial phonemes of the word targets in the word-monitoring condition. Word-initial rather than word-medial phoneme targets were used because an overwhelming number of phoneme-monitoring studies have adopted this as standard methodology. A marking tone, inaudible to subjects, was placed on the second track of the audiotapes at the beginning of each word-initial target phoneme (or target word). The tone started a timer that stopped when subjects pressed a response button. Timing and data collection were controlled by a PDP 11/05 computer.

Response booklets for measuring comprehension of the stories were prepared for each tape. The booklets contained a varying number of multiple-choice questions keyed to each story. The order of pages was determined by the presentation schedule of the stories on a given audiotape. Performance on these posttest questions was used to provide an objective measure of listening comprehension. There were a total of 48 questions. Twenty questions were factual in nature, requiring nothing more than recall of some explicitly stated information contained in the story; the remaining 28 questions required listeners to understand the implications of ideas and propositions expressed in the passages and to integrate these ideas with general knowledge.

Subjects

Forty-two naive students at Indiana University in Bloomington served as paid subjects in this study. They were recruited by means of an advertisement, and each reported no history of a hearing or speech disorder at the time of testing. The subjects were all right-handed, native speakers of English. Fourteen subjects were assigned to each tape set.

Procedure

Subjects were tested in groups of one to five. Each subject was seated in a booth out of direct sight of the others in a small testing room used for speech perception experiments.

Prerecorded instructions were presented at the beginning of each tape. The instructions and stories were presented binaurally over TDH-39 headphones. A typed copy of the instructions was placed at the front of each booklet to allow the subjects to read along as the instructions were read aloud. Typed copies of the stories were *not* available to the subjects.

The instructions emphasized that the primary concern of the experiment was to study how listeners understand and remember spoken stories. The subjects were told they would hear several short stories about a wide variety of topics and that their task was to answer the multiple-choice questions keyed to each story. They were told to do as well as they could, based on the information contained in the story they heard.

The subjects were also told that, for some stories, they would also be asked to listen for a particular target; either a word target or a word-initial sound target would be specified before the start of each story. In these cases, each subject was required to press a response button in front of him whenever he detected the presence of a particular target. The instructions emphasized that it was important to listen for the target throughout the entire story because it would occur several times. Speed and accuracy of responding were also stressed. The subjects were explicitly told, however, not to let the task of listening for a target interfere with their attempts to understand the story, because they would still have to answer comprehension questions about stories with targets in them.

The subjects were presented with the test stories in a self-paced format. The experimenter was present in the testing room and operated the tape recorder via remote control. Each story was presented only once for listening. After each story, the subjects immediately turned their booklets to the appropriate set of test questions and answered them by circling one of several response choices in pencil.

The subjects heard three practice stories, one from each monitoring condition, before actual testing began. They answered two comprehension questions for each practice story. After the experimenter answered questions clarifying the procedures and instructions, the test stories were presented. The entire experiment lasted about 45 min.

RESULTS

Mean latencies for the phoneme- and word-monitoring conditions were computed for each subject. These means were based on all reaction times that were longer than 100 msec and shorter than 1,600 msec. Reaction times outside this range were presumed to reflect anticipation, momentary inattention, or some other type of unusual processing strategy on the part of the listener. The overall means for the phoneme- and word-monitoring conditions are shown in Table 2. The total number of missed targets for these conditions was also computed for each subject. Table 2 presents the mean number of misses for both monitoring conditions.

As shown in Table 2, monitoring latencies are shorter for detecting words than phonemes. The observed pattern of reaction times is consistent with other reported findings of shorter latencies to word targets than to phoneme targets (see, e.g., Foss & Swinney, 1973; Savin & Bever, 1970). The results of at test for matched samples showed that the 93-msec difference between the two conditions was highly significant [$t(41) = 4.68$, $p < .001$].[1] The difference between the mean number of misses for word and phoneme targets, although small, was also reliable by a t test for independent samples [$t(82) = 2.91$, $p < .01$].

The multiple-choice comprehension questions for each of the 12 stories were scored separately for each subject. A composite error score was then obtained by accumulating, across subjects, the individual error scores for all the stories within each monitoring condition. This value was then expressed as a percentage of the total possible errors. The overall error scores for the three monitoring conditions were: word monitoring, 27.3%; phoneme monitoring, 30.5%; control, 26.5%. These data are shown in Panel A of Figure 1. None of the planned comparisons using independent t tests resulted in significant differences among the three conditions.

Panel B of Figure 1 presents comprehension performance for each monitoring condition, broken down in terms of errors on inferential and factual questions. For inferential questions, error scores were: word monitoring, 30.8%; phoneme monitoring, 31.1%; control, 28.8%. For factual questions, error scores were: word monitoring, 22.5%; phoneme monitoring, 25.7%; control, 21.1%. None of the error scores on the inferential questions were significantly different from one another. This was also true for performance on the factual questions.

DISCUSSION

Overall performance on the comprehension questions suggests that conscious focusing of a listener's attention on properties of the sound structure of the speech signal does not adversely affect spoken language understanding (at least under the conditions examined in the present experiment). Comprehension questions were responded to at similar levels in the word-, phoneme-, and nonmonitoring conditions. It is noteworthy that the level of comprehension performance observed in this experiment, about 70% correct, is similar to that obtained in a recent listening comprehension study reported by Pisoni (Note 2), using synthetic speech produced by rule. The approximate 30% error rate indicates that the level of difficulty of these stories was relatively high. Such performance levels reduce the possibility that the comprehension task was so easy for listeners that the expected Monitoring by Comprehension interaction would not be observed due to the process of ceiling effects.[2]

The finding of shorter latencies to words than to phoneme targets is an important one because it provides evidence that, in this study, subjects performed the monitoring tasks in ways analogous to subjects' performances in previous monitoring studies. Note that this

[1]The interested reader is referred to Foss, Harwood, and Blank (1980) for a recent discussion of previous interpretations ascribed to this reaction-time difference. In particular, Foss et al. discuss why it is a mistake to assume the following: (1) that the order of reaction times obtained in monitoring experiments reflects the order in which perceptual entities (like phonemes and words) are derived by listeners, and (2) that the perceptual entities that are derived the earliest are the primary units for speech processing. These two theoretically naive assumptions have led some theorists to conclude, apparently prematurely, that monitoring latencies are shorter for words than for phonemes because the former, *not* the latter, are the basic units of speech perception (see, e.g., Marslen-Wilson & Welch, 1978; Savin & Bever, 1970; Warren, 1971).

[2]Were we to have increased the difficulty of the subsidiary monitoring task by requiring detection of several different target items, we might have succeeded in affecting comprehension performance (see, e.g., Logan, 1979). Such a finding, while interesting in and of itself, would not weaken any of the conclusions drawn from the present study. After all, the standard methodology used in monitoring experiments investigating fluent speech processing does not involve listening for several different target items.

result was observed in spite of the fact that, for a given story, the phoneme target always appeared within the same word. While the word containing the phoneme target was not disclosed to the subject at the start of the story, the word containing the target phoneme occurred more than six times in the average passage. Thus, after the first target was identified, subjects could have switched to a word-monitoring strategy. If this had happened, the observed pattern of response times should not have replicated earlier findings of RT differences between the two monitoring conditions.

The absence of a significant Monitoring by Comprehension interaction in this study is important because critics of on-line monitoring paradigms as measures of listening comprehension have assumed, without empirical support, that conscious attention to the sound structure of an utterance at any level interferes with normal comprehension. Since the existence of a Monitoring by Comprehension interaction was an intuitively sound prediction, it is quite appropriate to examine any, and all, reasonable alternative interpretations of the present findings.

Perhaps subjects chose to sacrifice performance on the monitoring task to insure the availability of processing resources for comprehension. This possibility seems ruled out on several accounts. First, as already pointed out, the pattern of monitoring latencies reported here replicates earlier findings of shorter RTs to word targets than to phoneme targets. Second, absolute RTs for word and phoneme targets are also quite similar to earlier studies (about 500 msec). Third, the number of detection misses obtained for each target type is in reasonable accord with other monitoring studies (about 4%). Fourth, as in other monitoring studies, there is no evidence of a speed-accuracy tradeoff.

The reasonable conclusion to draw from the present data, then, is that listening comprehension does not appear to be impaired by simultaneous monitoring of single words or phonemes. This is an unexpected result, since it undermines the specific criticism that on-line monitoring paradigms of this kind interfere with normal speech processing. More specifically, phoneme- and word-monitoring data may not be neglected on the grounds that the tasks interfere with normal comprehension.

In the introduction, we asked whether monitoring at the word level would have the same effects on comprehension as would monitoring at the phonemic level. We hoped that the answer to this question would provide some insights into the respective roles of lexical and phonemic representations in speech processing. The present findings do not support the view that only lexical representation plays a critical role in understanding spoken language (as Marslen-Wilson & Tyler, 1980, have claimed). This view would predict that word and phoneme monitoring should have differential effects on comprehension performance; our results show they do not. The finding of equivalent effects is consistent with the view that both phonemic and lexical representations are normally computed during fluent speech processing, and that both are important for spoken language understanding (see, e.g., Foss & Blank, 1980). We should, however, point out that misses for phoneme targets were slightly higher than misses observed for word targets. While this may be indicative of some underlying processing difference, it may also simply reflect differences in subjects' familiarity with the concepts of words and phonemes. Whatever the cause, the differential miss rates are not associated with differences in comprehension performance. Thus, we find little evidence to support the contention that one level (lexical) is more primary or central than another (phonemic) to *understanding* fluent speech.

The present results indicate that the ultimate level of understanding achieved for the content of a spoken message is unaffected by an ancillary monitoring task. The present results do not, however, shed light on whether ongoing comprehension processes are affected by the

task demands of word or phoneme monitoring. It may be, for example, that monitoring slows down the speed at which listeners can execute the processes involved in comprehension, while leaving the end product (namely, understanding) intact. An empirical test of this requires collecting latencies to comprehension questions as well as error rates. Experiments of this kind are currently underway in our laboratory.

In summary, the results of the present study suggest that spoken language understanding is not adversely affected by simultaneous, conscious monitoring of phonemic or lexical properties of the speech signal. Subjects' understanding of spoken stories, as measured by performance on multiple-choice comprehension questions, did not differ across phoneme-, word-, and no-monitoring conditions. This result calls into question the specific claim that target-monitoring tasks selectively interfere with understanding fluent speech.

Acknowledgments

This paper is based on research conducted at Indiana University in Bloomington. The work reported here was supported by NIH Training Grant NS-07134-01 and NIMH Research Grant MH-24027-06. We thank Jerry C. Forshee and David Link for their technical help and Robert E. Remez for comments on an earlier draft of the paper.

References

Aaronson D. Performance theories of sentence coding: Some qualitative observations. Journal of Experimental Psychology: Human Perception and Performance. 1976; 2:42–55.

Ammon PR, Ostrowski B, Alward K. Effects of task on the perceptual organization of sentences. Perception & Psychophysics. 1971; 10:361–363.

Blank MA. Measuring lexical access during sentence processing. Perception & Psychophysics. 1980; 28:1–8. [PubMed: 7413405]

Carey PW. Verbal retention after shadowing and after listening. Perception & Psychophysics. 1971; 9:79–83.

Cole, RA.; Jakimik, J. A model of speech perception. In: Cole, RA., editor. Perception and production of fluent speech. Hillsdale, N.J: Erlbaum; 1980.

Foss DJ, Blank MA. Identifying the speech codes. Cognitive Psychology. 1980; 12:1–31. [PubMed: 7351123]

Foss, DJ.; Harwood, DA.; Blank, MA. Deciphering decoding decisions: Data and devices. In: Cole, RA., editor. Perception and production of fluent speech. Hillsdale, N.J: Erlbaum; 1980.

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.

Klatt DH. Speech perception: A model of acoustic-phonetic analysis and lexical access. Journal of Phonetics. 1979; 7:279–312.

Levelt, WJM. A survey of studies in sentence perception: 1970–1976. In: Levelt, WJM.; Flores d'Arcais, GB., editors. Studies in the perception of language. New York: Wiley; 1978.

Logan GD. On the use of a concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology: Human Perception and Performance. 1979; 5:189–207.

Marslen-Wilson WD, Tyler LK. The temporal structure of spoken language understanding. Cognition. 1980; 8:1–71. [PubMed: 7363578]

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 of phoneme identification. Journal of Verbal Learning and Verbal Behavior. 1976; 15:43–52.

Rudnicky, A. Structure and familiarity in the organization of speech perception. Carnegie-Mellon University; Pittsburgh, Pennsylvania: 1980. Unpublished dissertation thesis

Savin HB, Bever TG. The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior. 1970; 9:295–302.

Warren RM. Identification times for phonemic components of graded complexity and for spelling of speech. Perception & Psychophysics. 1971; 9:345–349.

Warren, R. Auditory illusions and perceptual processes. In: Lass, NJ., editor. Contemporary issues in experimental phonetics. New York: Academic Press; 1976.

REFERENCE NOTES

1. Blank, MA. Decoding fluent speech: The role of sensory- and knowledge-based processes. 1980. Manuscript submitted for publication
2. Pisoni, DB. Research on Speech Perception Progress Report No. 5. Bloomington: Indiana University; 1979. Some measures of intelligibility and comprehension.

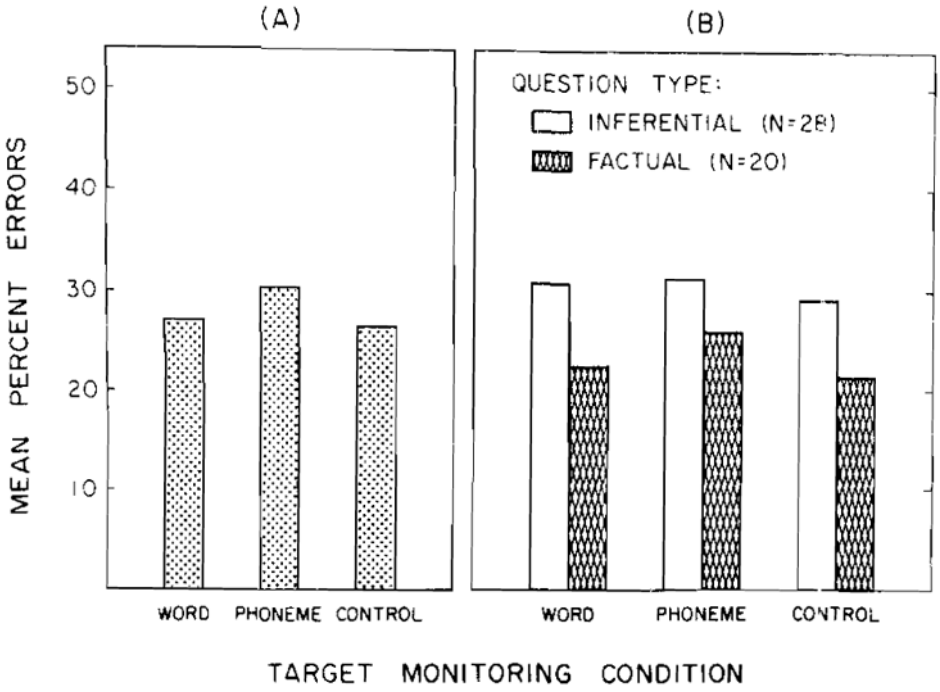

Figure 1.
Error rates for comprehension questions as a function of monitoring condition. Overall performance is shown in Panel A; performance broken down in terms of factual and inferential questions is shown in Panel B.

Table 1

Description of Stories and Questions Used to Measure Comprehension

						Number of Comprehension Questions		
Story Number	**General Topic**	**Number of Words**	**Target Phoneme**	**Target Word**	**Target Frequency**	**F**	**I**	**T**
1	Measuring Star Distances	161	/d/	distance	4	1	2	3
2	Verbal Communication	315	/g/	group	5	1	5	6
3	An Outdoor Scene	278	/g/	green	3	2	1	3
4	Insufficient Oxygen	363	/b/	behavior	8	1	4	5
5	Aluminum	273	/m/	metal	5	3	3	6
6	Carbon Dating	374	/k/	carbon	14	2	3	5
7	The American Party System	490	/b/	both	5	2	3	5
8	Effective Communication	343	/t/	talk	12	3	1	4
9	Basketball	270	/m/	men	4	1	2	3
10	Science Fiction	271	/b/	books	5	2	0	2
11	Self-Actualization	228	/g/	goal	2	1	2	3
12	Hair Today; Gone Tomorrow	313	/b/	bald	10	1	2	3

Note–F= factual; I = inferential; T= total.

Percept Psychophys. Author manuscript; available in PMC 2012 November 28.

Table 2

Mean Latencies (in Milliseconds) and Mean Number of Misses for Word- and Phoneme-Monitoring Conditions

	Mean Latency	Mean Misses
Word Monitoring	574	4.5
Phoneme Monitoring	667	7.5