



Published in final edited form as:

*J Verbal Learning Verbal Behav.* 1983 December ; 22(6): 609–632. doi:10.1016/S0022-5371(83)90365-1.

## Cracking the Dual Code: Toward a Unitary Model of Phoneme Identification

Donald J. Foss and Morton Ann Gernsbacher

University of Texas at Austin

### Abstract

The results of five experiments on the nature of the speech code and on the role of sentence context on speech processing are reported. The first three studies test predictions from the dual code model of phoneme identification (Foss, D. J., & Blank, M. A. *Cognitive Psychology*, 1980, **12**, 1–31). According to that model, subjects in a phoneme monitoring experiment respond to a prelexical code when engaged in a relatively easy task, and to a postlexical code when the task is difficult. The experiments controlled ease of processing either by giving subjects multiple targets for which to monitor or by preceding the target with a similar-sounding phoneme that draws false alarms. The predictions from the model were not sustained. Furthermore, evidence for a paradoxical nonword superiority effect was observed. In Experiment IV reaction times (RTs) to all possible /d/-initial CVCs were gathered. RTs were unaffected by the target item's status as a word or nonword, but they were affected by the internal phonetic structure of the target-bearing item. Vowel duration correlated highly (0.627) with RTs. Experiment V examined previous work purporting to demonstrate that semantic predictability affects how the speech code is processed, in particular that semantic predictability leads to responses based upon a postlexical code. That study found “predictability” effects when words occurred in isolation; further, it found that vowel duration and other phonetic factors can account parsimoniously for the existing results. These factors also account for the apparent nonword superiority effects observed earlier. Implications of the present work for theoretical models that stress the interaction between semantic context and speech processing are discussed, as are implications for use of the phoneme monitoring task.

---

When listeners process spoken language they develop a series of representations of the input which range in content from the low level physical structure of the speech signal to a high level semantic representation. A theory of comprehension must specify the nature of those representations or codes, and must detail the mechanisms by which they are developed. Here we will be concerned with two subproblems in the general area of speech perception. One question that we will address, one that has received considerable attention, is whether the listener computes a representation of the input whose units are relatively low level phonetic segments. Our second question is whether and to what extent the processing of low level codes is affected by the ongoing syntactic and semantic analyses of the input, that is, whether the codes truly interact in some way.

With respect to the first question, an earlier paper (Foss & Blank, 1980, p. 2) identified two broad classes of models: “On the one hand are models in which phonetic segments are computed during the transformation of the acoustic signal into a lexical representation. On the other hand are models in which no such segments are computed.” Foss and Blank concluded that the former class of models was the correct one.

Some of the evidence they reviewed, and all of the new evidence that they presented, was gathered via the phoneme monitoring technique—a technique in which subjects are asked to listen to and comprehend speech and to respond by pushing a button when a specified target phoneme (usually word-initial) is detected. One issue that arises when using the phoneme monitoring technique concerns the point during processing at which the listener responds to the target. This methodological issue is tied closely to the first of the aforementioned theoretical ones. If responses to target phonemes can be made prior to the time at which the word is identified, then this is strong evidence for the class of models claiming that listeners compute a phonetic representation of the speech signal en route to lexical access. Considerable evidence, though not all of it, is consistent with just this view.

Foss and Blank presented evidence that listeners can respond to a word-initial target phoneme prior to accessing the word it begins. They showed, for example, that time to respond to initial phonemes was just as rapid when the word carrying the target has a low frequency of occurrence in the language as when it has a high frequency; likewise, subjects are just as fast to respond to a target which begins a pronounceable, nonlexical item (e.g., a nonword) as to one which begins a real word. These findings suggest strongly that listeners need not access the target-bearing item before responding. If such access was required, then reaction times (RTs) to the target phoneme would reflect differences in time to access the target-bearing word (or nonword). RTs would be slower when targets occurred on low frequency words or on nonwords as compared to when they occurred on high frequency words.

Evidence that subjects sometimes respond to targets subsequent to lexical access was also reviewed by Foss and Blank. One important study that supports this view, a study that we will have occasion to examine more carefully later, was conducted by Morton and Long (1976). Their evidence also bears directly on the second of Our theoretical issues, the question of whether the mechanisms that compute the low level codes are affected by the surrounding syntactic and semantic context. In Morton and Long's experiment subject were presented with sentences like (1) and (2) and were asked to respond to a target phoneme (the target is /b/ in these examples).

- (1) He sat reading a *book* until it was time to go home for his tea.
- (2) He sat reading a *bill* until it was time to go home for his tea.

Morton and Long found that subjects responded more rapidly when the target phoneme was carried by a word that was contextually more probable (e.g., *book*) than when it began a less likely word (e.g., *bill*). This result was replicated by Foss and Blank (1980) and by Dell and Newman (1980), though in a modified form and with somewhat different results in the latter case. Morton and Long interpreted their result as showing that listeners can use the context

to identify rapidly the target-bearing word and that they can then identify and respond to its word-initial target phoneme. In another study, Blank (Note 1) found that subjects responded to a word-initial target more rapidly when the target-bearing word (an object noun) was semantically related to the main verb than when it was not so related. Thus, RTs to a /b/ target on *baby* were shorter when the preceding verb was *diapered* than when it was *comforted*. In addition, a further study consistent with the view that listeners sometimes respond to word-initial targets subsequent to word identification—and that higher level processes affect those at lower levels—was carried out by Rubin, Turvey, and Van Gelder (1976). They presented subjects with lists comprised of words and nonwords, and asked them to carry out the phoneme monitoring task. RTs were shorter when target phonemes occurred on words than when they occurred on nonwords.

Foss and Blank proposed a “dual code” model to account for the various sets of data obtained via use of the phoneme monitoring task. According to this model, subjects are able to gain access and respond to codes both at a prelexical or phonetic level (i.e., prior to the point at which the word carrying the target phoneme has been identified) and at a postlexical or phonological level (i.e., after the word carrying the target has been identified). In the former case, according to the hypothesis, subjects examine the acoustic/phonetic representation of the incoming signal and respond when it matches the target specification. In the latter case, subjects examine the phonological information stored with the word in memory after they recognize that word. A response is made when the subjects discover that the word begins with the specified target phoneme. According to the dual code model, responding to the prelexical phonetic code is relatively difficult and subjects are able to do so only under certain conditions. For example, one important condition is that the signal-to-noise ratio be high enough that the acoustic representation of the input actually carries the relevant information. Another is that the subjects are able to devote a substantial fraction of their processing resources to the monitoring task. The latter is required since the model assumes that the low level phonetic code fades or is overwritten rapidly; unless the subject can closely monitor the representation at this level, the target will not usually be detected. Thus, the model predicts that subjects will respond to the postlexical code both when other processing demands preclude constant examination of the prelexical representation and when the signal-to-noise ratio is low.

In this paper we present a number of tests of the dual code model. These tests rest upon the assumption that one can manipulate the probability of subjects' responding to the prelexical code by manipulating the processing demands made upon the listeners while they are carrying out the monitoring task. In addition, we will show that some studies in the literature purporting to show effects of semantic context on phoneme detection are missing important controls, and that the conclusions about speech processing that one may draw from them may change greatly when these controls are added. Also, and importantly, we will show that RT data gathered with the phoneme monitoring task fit nicely with some important data and hypotheses drawn from the speech perception literature. To begin, we will test two predictions from the dual code model and find them wanting in particular ways. We also discover, and then explain, a paradoxical “nonword superiority effect.” We show that listeners carrying out the phoneme monitoring task are greatly affected by the phonetic

nature of the target stimuli, while appearing not to be affected by the semantic context in which the target word occurs. Finally, a revised model of phoneme detection will be offered, and implications for speech processing will be discussed.

## Experiment I: Effects of Multiple Targets on Reaction Times

According to the dual code model, listeners must be able to examine frequently the prelexical representation of speech if they are going to respond to the phonetic code. This is required since such a representation is quickly replaced by the representation of subsequent speech input. When subjects cannot rapidly examine the phonetic representation they must respond instead to the postlexical phonological code. It follows, then, that one could manipulate the probability that subjects will respond to one or the other of these codes if one could control the rate or effectiveness of examining the prelexical code. The first experiment was designed to affect that rate by giving subjects a task that was demanding of their processing resources.

In this experiment, subjects were asked to monitor simultaneously for two word-initial target phonemes and to press a button when they detected either one of them. The requirement to monitor for two targets should reduce the probability that subjects can detect the target phoneme at the prelexical level since continually examining the input for both targets will take time and resources. On many trials the target might not be detected before the prelexical representation has been overwritten by new input. In those cases the subject must respond at the postlexical level if he or she is to respond at all.

In this experiment subjects were presented with sentences such as (3)–(6). These sentences are identical save for the occurrence of nonwords for words and the substitution of items beginning with /d/ for those beginning with /b/.

- (3) Mary was often rewarded with a *bowl* of ice cream after eating all her vegetables.
- (4) Mary was often rewarded with a *bap* of ice cream after eating all her vegetables.
- (5) Mary was often rewarded with a *dish* of ice cream after eating all her vegetables.
- (6) Mary was often rewarded with a *dap* of ice cream after eating all her vegetables.

Subjects were asked to comprehend these sentences and to respond by pushing a button whenever a word began with either of two possible targets, /b/ or /d/.

Earlier work (Foss & Blank, 1980) showed clearly that subjects who are asked to monitor for a single phoneme respond to targets that begin nonwords just as rapidly as they do to targets beginning real words. Indeed, that result provided key evidence for the claim that listeners can respond prior to accessing or identifying the target-bearing item—the word or nonword. (The same work also showed that targets occurring immediately *after* a nonword are responded to significantly more slowly than those occurring right after a real word; thus, the status of an item as a word or nonword does affect phoneme monitoring RTs when targets are placed after the critical items.) If the requirement to monitor for two targets lowers the probability that subjects can respond at to the prelexical code, then the subjects in

this experiment will have a higher probability of responding to the postlexical code. This means that subjects will have gained access to the word (or nonword) carrying the target before a response is initiated. Therefore, RTs will reflect such access times.<sup>1</sup> Of course, the time needed to identify a word is likely to be much less than the time needed to construct a representation for a nonword and to examine it. Thus, the listener will be able to examine and respond to the postlexical code associated with the word much more rapidly than the code associated with the nonword. The average RTs to respond to the word-initial targets should therefore be significantly less than the RTs to respond to the nonword-initial targets.

The present study should, then, yield results that have a different pattern from those observed by Foss and Blank. They did not observe any effect due to the status of the target bearing item (word vs nonword). Here we expect that RTs will differ significantly depending upon the status of the target-bearing item, with longer RTs being observed when the target is on a nonword.

## Method

**Design and materials**—Thirty-two basic experimental sentences were constructed. Each sentence had four versions as in examples (3)–(6): A sentence contained either a single nonsense word that began with the phoneme /b/, a single nonsense word that began with the phoneme /d/, a real English word that began with the phoneme /b/, or a real English word that began with the phoneme /d/. In order that each basic sentence could occur in each condition across the experiment, four material sets were constructed. Each material set contained all 32 basic sentences; 8 sentences in each material set came from each of the four conditions. Across the material sets, each basic sentence occurred in all four conditions. The experiment was, therefore, a 2 (word type: real/nonsense word) by 2 (target phoneme: /b/ vs /d/) by 4 (material sets) factorial, with the first two variables within subjects and the last variable between subjects.

The target-bearing words used in this experiment were equated, across /b/-initial and /d/-initial words, for frequency, syllable length, and syllabic stress. The mean Kuera and Francis (1967) frequency of the /b/-initial words was 52.87 and the mean frequency of the /d/-initial words was 58.71. The mean length in syllables of both the /b/-initial and /d/-initial words was 1.41. Both sets of target-bearing words carried their primary stress in the initial syllable. The /b/-initial and /d/-initial words were comparable in meaning (e.g., bowl/dish; Boston/Dallas; bishops/deacons) and were selected so that the overall sentential meaning would be highly similar.

Two nonsense words were constructed for each experimental sentence, one beginning with /b/, the other with /d/. Both nonsense words were the same length as the two real words that they replaced, and they rhymed with each other. A few examples of real/nonsense word quadruplets are bowl, bap, dish, dap; Boston, Beggel, Dallas, Deggel; bishops, beppems, deacons, deppems.

---

<sup>1</sup>Of course, subjects cannot access a preexisting representation of a nonword since there is not one to access. In the case of nonword targets in this experiment we conjecture that listeners construct a representation of its phonological “spelling” and respond when that representation satisfies the target description. Such a construction process will generally be more time consuming than the one required to access a preexisting phonological spelling for a real word.

Thirty-two filler sentences were constructed. Sixteen of them did not contain a target phoneme; eight of these contained a nonsense word and eight did not. Another 16 filler sentences contained one of the two target phonemes. Eight of these began with the target /b/ and eight began with the target /d/, half of each eight being a real word and the other half a nonsense word. In addition, half of the filler words beginning with a target were adjectives and half were adverbs. Balanced across these two word classes, half of the filler targets occurred relatively early in the sentences and half occurred relatively late. The 64 sentences were randomized with the constraints that in each of the 8 blocks of 8 sentences, 4 target sentences appeared, one of each experimental type; two filler sentences without targets appeared, one with a nonsense word, one without; and two filler sentences with targets appeared, with the target phoneme beginning either an early/late, adjective/adverb, nonsense/real word. No more than two experimental or filler sentences occurred consecutively.

A female speaker recorded each of the four material sets on one channel of a tape. A pulse, inaudible to subjects, was placed on the second channel of the tape at the beginning of the target phoneme. The pulse started a timer which recorded reaction times for subjects to press their response buttons.

**Subjects:** The subjects were 48 undergraduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Twelve subjects were randomly assigned to each of the four experimental tapes (material sets).

**Procedure:** Subjects were tested in groups of one to six, with the experimenter and subjects occupying adjoining rooms. Each subject was seated in a booth out of direct sight of the others. Instructions outlining the task were recorded at the beginning of each experimental tape, which was presented binaurally over headphones. Subjects were told to lightly rest the index finger of their preferred hand on the response button in front of them. They were also told to (a) comprehend each sentence, and (b) push the response button whenever they heard a word that started with a /b/ or a /d/. A trial consisted of the word “ready,” about 2 seconds of silence, and the presentation of the sentence. Subjects were told that some sentences would contain a nonsense word and that others would contain only real English. They were instructed not to let this interfere with their monitoring and comprehension tasks. Subjects were informed that the occurrence or nonoccurrence of the target sound was not determined by the presence or absence of a nonsense word in the sentence. They were pre-warned that after some of the sentences the speaker on the tape would say “paraphrase,” and that they were then to write a paraphrase of the sentence they just heard. This instruction emphasized the importance of paying close attention to the meaning of the sentences. Pencil and ruled paper were supplied in each of the subject's booths. Subjects were allowed 45 seconds for the paraphrase task. Sixteen of the filler sentences, eight with and eight without target phonemes, were tested in this manner. After writing their paraphrase, subjects were told to replace their finger on the response button.

Following the instructions, subjects were given four practice sentences, one of which did not contain a target, and an additional practice sentence that they were asked to paraphrase.

After the experimenter answered questions clarifying any uncertainties regarding the instructions, the experimental and filler sentences were presented.

## Results

In all of the experiments reported in this paper the RT data have been truncated in the following way. A mean and standard deviation was computed for each subject and for each item in the experiment. If any individual RT was more than two standard deviations from both the mean for the subject and the mean for the item, it was omitted and replaced, following a procedure suggested by Winer (1971). Missing data points were also filled according to Winer's procedure.

The results of this experiment are shown in Table 1. Analyses of variance performed on the mean correct reaction time data, by subject and by item, revealed a main effect of target phoneme. RTs to target phoneme /b/ ( $\bar{X} = 490$  milliseconds) were much shorter than to target phoneme /d/ ( $\bar{X} = 546$  milliseconds),  $\min F(1,59) = 8.78, p < .01$ . There was no main effect due to the lexical status of the target bearing item,  $F/1 < 1; F/2 < 1$ ; however, there was a significant interaction between phoneme target and lexical status such that /b/ words ( $\bar{X} = 471$  milliseconds) were responded to much more rapidly than /b/ nonwords ( $\bar{X} = 508$  milliseconds) yet /d/ words ( $\bar{X} = 569$  milliseconds) were responded to much more slowly than /d/ nonwords ( $\bar{X} = 523$  milliseconds),  $\min F(1,50) = 12.88, p < .001$ . The pattern of this interaction was such that the main effect for lexical status disappeared when data from the /b/ and /d/ items were analyzed jointly. Therefore, planned comparisons, that is, analyses of variance, were performed on the data from the /b/ items only and the /d/ items only. Both additional analyses revealed main effects of lexical status (they were, of course, in opposite directions), /b/ items only,  $\min F(1,51) = 5.19, p < .05$ ; /d/ items only,  $\min F(1,45) = 5.06, p < .05$ . No other main effects or interactions were significant in any of the original analyses or planned comparisons.

## Discussion

The dual code model suggests strongly that RTs to targets on nonwords should be longer than those on words when subjects are given a difficult task such as monitoring for multiple targets. The model predicts in such a case that subjects should not be able to attend and respond to the prelexical phonetic code because their processing resources will be too taxed to do so. Instead, according to the model, they will respond to the postlexical code, and thus variables that affect speed of lexical access will also affect RTs. The results of Experiment I were therefore quite confusing when looked at from the perspective of the dual code model. On the one hand, RTs were faster when targets were on words rather than on nonwords, just as the model predicted—but this only held when the target phoneme was /b/. (In pilot work we had examined RTs to /b/ targets and obtained these expected results, so this part of the experiment confirmed that pilot work.) On the other hand, when the target was /d/, we observed a very paradoxical result: a non-word superiority effect. RTs were faster when the target was on a nonword than when it was on a word.

Of course, a nonword superiority effect is not only paradoxical when looked at from the vantage point of the dual code model. It is an untoward finding from nearly any perspective,

and it begs to be explained away. One possible explanation, one that is consistent with the dual code model, is based on the fact that items with initial /d/'s lead to considerably more confusion about word boundaries than /b/ items. A concern with word boundary phenomena is important since subjects in this phoneme monitoring experiment were told to respond only to word initial /b/'s or /d/'s. Medial or final /b/'s or /d/'s were not targets. Thus, the listeners must have determined that a word boundary occurred just prior to the /b/ or /d/ before they responded.

It is a fact of English that many more words end in /d/ than end with /b/. For example, of the 800 most frequently occurring words according to Kucera and Francis (1967), two of them end in /b/ while 75 of them end in /d/. Also, in all the materials used in Experiment I, no word ended in /b/ while 73 ended in /d/. And in a randomly selected page from an introductory psychology textbook, no words ended in /b/ while 50 ended in /d/. In all of these sources the proportion of words beginning with /b/ and /d/ was approximately the same.

Consider, then, the decision faced by a subject in this task. When a /b/ occurred, the subject could be quite certain that it was a target since the majority of /b/'s occurred in initial position. When a /d/ occurred, however, the subject could not be sure that it was word initial; quite often it was not. Thus, it is likely that a greater segmentation problem existed with /d/ targets than with /b/ targets. This may account for at least part of the main effect due to phoneme; average RT to /d/ targets was significantly longer than to /b/ targets. Although perhaps less likely, it is conceivable that the segmentation problem was inadvertently made more difficult in the case of the /d/ words than in the case of the /d/ nonwords. If so, then RTs to respond to /d/-word targets would be slowed. While there is, admittedly, little reason a priori to credit this hypothesis, we conducted a test of it in light of the fact that we do not know very much about the processes of segmentation. Indeed, this is one of the gaps that work on speech perception in natural contexts such as the present effort is meant to fill.

In order, then, to investigate the apparent nonword superiority effect with the /d/ targets, a control experiment was conducted.<sup>2</sup> Its main purpose was to determine whether the results of Experiment I were due to differences in segmentation difficulty experienced by those subjects. The logic of the control study was as follows. Let us assume first that the results of Experiment I were affected by differences in segmentation difficulty between the /b/ items and the /d/ items. In that case, the pattern of results in a list experiment should change dramatically since the need for segmentation does not arise in auditorily discrete lists; there, of course, the items are already clearly segmented. On the other hand, if we assume that the results of Experiment I were not due to segmentation problems but to some other source, then we might again see the anomalous nonword superiority effect when /d/ targets occur in lists.

The 16 graduate student subjects in the control study were presented with a 360 item list and were asked to respond by pushing one button whenever an item began with a /b/ and by pushing another button whenever one began with a /d/. Half of the 64 target items within

---

<sup>2</sup>Details of this test may be obtained from the authors.



this list began with each of these phonemes. Also, half of the targets began words and half began non-words. In fact, the target items on this list were the target-bearing words and non-words from Experiment I. They were clearly segregated since the ISI, though of random duration, was always greater than 1 second. To clarify, if the results of Experiment I were due to segmentation problems, then we should observe a word superiority effect for both the /b/ and the /d/ targets, essentially replicating with different materials the results of Rubin et al. (1976) described earlier. Any other pattern of results would suggest strongly that the anomalous results of Experiment I were not due to segmentation problems.

The results of the control experiment were quite clear. As previously observed in Experiment I, there was no main effect of lexical status on reaction times,  $F(1,15) = 1.08$ , but there was the paradoxical phoneme target by lexical status interaction. Again, /b/ words ( $\bar{X} = 593$  milliseconds) were responded to more rapidly than /b/ nonwords ( $\bar{X} = 600$  milliseconds), whereas /d/ words ( $\bar{X} = 585$  milliseconds) were responded to more slowly than /d/ nonwords ( $\bar{X} = 566$  milliseconds),  $F(1,15) = 6.78, p < .05$ . Thus, in important respects, the results of this control study replicated the pattern observed in Experiment I. Subjects responded to the target /b/ faster when it began a word than when it began a nonword (though this difference taken by itself was not significant in the list study); and they responded to the target /d/ significantly more slowly when it began a word than a nonword. In short, we again observed the paradoxical nonword superiority effect when the target was /d/. This permits us to infer that the nonword superiority effect observed for /d/ targets in Experiment I was not due to problems of segmentation associated with the /d/ targets. The solution to the paradox must reside elsewhere. (We will note in passing that RTs to /d/ targets were somewhat faster than to /b/ targets here, though they were slower in Experiment I. This suggests that the main effect due to target in the first study was due to word boundary problems associated with /d/ targets.)

At this point, then, we are in the position of having observed data in Experiment I that do not confirm a prediction from the dual code hypothesis. Recall that we anticipated that a word superiority effect would be observed when subjects monitored for multiple targets on words and nonwords. No such effect obtained. We will leave hanging for the moment both of the issues at hand — the paradoxical nonword superiority effect with /d/ targets, and the failure of a prediction based on the dual code model. Both of these matters will, of course, be taken up again shortly. In the introduction we noted that we would present a variety of tests of the dual code model. We will now turn our attention to another such test.

## Phonetic Similarity

Dell and Newman (1980) proposed a view of phoneme monitoring similar in some respects to the dual code model, and they conducted a very interesting test that supported this view. Subjects in their experiments were presented with sentences like those used by Morton and Long (1976) in which the transitional probability or semantic predictability of the target-bearing words was manipulated — as in the sentence The surfers drove to a beach/bay to try out the waves. However, Dell and Newman preceded the target-bearing words with an adjective that itself began with a phoneme that was either phonetically similar to the target phoneme (e.g., /p/ as in *private*), or dissimilar to it (e.g., /s/ as in *secret*). They found that the

advantage accruing to predictable words was much greater when the targets were preceded by adjectives beginning with similar phonemes than when they began with dissimilar ones. This result fits the dual code model nicely. If we assume that subjects are likely to execute a false alarm to words beginning with phonemes similar to the target, and that they are more likely to devote processing resources to those foils than to dissimilar ones, then the necessary resources may not be available to inspect the prelexical phonetic code when the actual target occurs. In that case, then, subjects will respond to the postlexical code when it becomes available. And it is reasonable to assume that it will become available earlier for the predictable word than for the unpredictable one.<sup>3</sup>

We extended the test of the dual code model by manipulating both the phonetic similarity of the word preceding the target item as well as the target-bearing items themselves—the latter were either words or nonwords.

## Experiment II: Effects of Phonetic Similarity

The logic of this experiment is straight-forward. Subjects were asked to monitor for target phonemes that occurred on either words or nonwords, as in Experiment I (though in the present study only a single target was specified). The target-bearing items were immediately preceded by words that began with a phoneme that was either phonetically similar to the target or dissimilar to it. As noted above, we assume that initial phonemes which are similar to the target will require more of the subject's resources than will dissimilar ones. In consequence, the probability that subjects can respond to the attention-demanding prelexical code will be less in the case of the similar-sounding foils. Therefore, since subjects will be likely to respond to a postlexical code in those instances, we should observe a difference between RTs to targets on words and nonwords, with the former being more rapid. While the prediction from the dual code model did not work out unambiguously in the case of the earlier multiple target experiment (Experiment I), the prediction seems more likely to be supported here because of the similar work published by Dell and Newman.

### Method

**Design and materials**—Thirty-two basic experimental sentences were constructed, each containing a target-bearing item. The target was always /b/. Each sentence had four versions: A sentence contained either a single nonsense word or only real English words; crossed with this variable, immediately preceding the nonsense/real target item was a word with either a highly similar initial phoneme /p/, or a dissimilar one /s/. This defines four conditions. In order that each basic sentence could occur in each condition across the experiment, four material sets were constructed. Each material set contained all 32 basic sentences; 8 sentences in each material set came from each of the four conditions. Across the material sets, each basic sentence occurred in all four conditions. The experiment was, therefore, a 2

---

<sup>3</sup>Dell and Newman also found no significant effect due to predictability when the target-bearing words were preceded by adjectives beginning with dissimilar phonemes. This result was not entirely expected given the data obtained by Morton and Long and by Foss and Blank. However, Dell and Newman noted that there were a number of differences between their experiment and those carried out by these other investigators, and they further pointed out that some of these differences might lead to subjects responding to the prelexical code in the case of dissimilar prior phonemes. In addition, Dell and Newman's model proposed that the pre- and postlexical codes are processed in parallel, and they gave an account of this result in terms of that model.

(word type: nonsense/real) by 2 (initial phoneme of preceding word: similar/dissimilar) by 4 (material sets) factorial, with the first two variables within subjects and the last between subjects.

The preceding adjectives were equated, across /p/-initial and /s/-initial words, for frequency and syllable length. The mean Ku era and Francis (1967) frequency of the /p/-initial adjectives was 72.86 and the mean frequency of the /s/-initial adjectives was 75.25. The mean length in syllables of both the /p/-initial adjectives and the /s/-initial adjectives was 1.81. The /p/-initial and /s/-initial adjectives were comparable in meaning (e.g., painful, stinging) and were selected so that the overall sentential meaning would be highly similar between the two conditions.

The nonsense words were derived from the real English words that they replaced. Each nonsense word shared with its counterpart the same initial phoneme /b/, syllabic structure, and word stress pattern. A few examples of the nonsense/real word pairs are bem/burn, bunnelled/buttered, broop/brass. Half (16) of the nonsense/real word pairs were nouns; the other half were verbs. None of the nonsense/real word pairs occurred within the first or last four words of an experimental sentence. An example experimental sentence is While her coffee was still warm, Polly/Susan bunnelled/buttered her toast with margarine.

Thirty-two filler sentences were constructed. Sixteen filler sentences did not have the target phoneme in them; 8 of these contained a nonsense word and 8 did not. Another 16 filler sentences contained the target phoneme. Eight of these contained a nonsense word and 8 did not. In addition half of the target-bearing filler words were adjectives and half were adverbs. Balanced across these two word classes, half of the target items on fillers occurred relatively early in the sentences and half occurred relatively late.

The 64 sentences were randomized with the constraints that in each of the 8 blocks of 8 sentences 4 target sentences appeared, one of each experimental type; two filler sentences without targets appeared, one with a nonsense word, one without; and two filler sentences with targets appeared. No more than two experimental or two filler sentences occurred consecutively. A male speaker recorded each of the four material sets on one channel of a tape. A pulse, inaudible to subjects, was placed on the second channel of the tape at the beginning of the target phoneme. The pulse started a timer which was stopped when subjects pressed a button.

**Subjects**—The subjects were 64 under-graduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Sixteen subjects were assigned to each of the four experimental tapes (material sets).

**Procedure**—The procedure was quite similar to that used in Experiment I, including use of the paraphrase task, with the exception that subjects were only given a single target /b/.

## Results

Results from Experiment II are shown in Table 2. Analyses of variance, by subject and by item, on these data showed that, overall, the effect of preceding phoneme was significant. As expected, RTs were substantially longer when the word preceding the target-bearing item began with the initial phoneme /p/ ( $\bar{X} = 571$  milliseconds) as opposed to /s/ ( $\bar{X} = 529$  milliseconds),  $\min F'(1,44) = 7.02, p < .001$ . A main effect of lexical status was also realized, such that RTs to target-bearing words ( $\bar{X} = 536$  milliseconds) were significantly shorter than those to target-bearing nonwords ( $\bar{X} = 564$  milliseconds),  $\min F'(1,63) = 5.51, p < .05$ .

The important interaction between preceding phoneme and lexical status reached significance in the analysis which treated subjects as a random effect,  $F/1(1,60) = 5.73, p < .05$ , but was only marginally significant in the analysis with items considered a random effect,  $F/2(1, 28) = 3.65, p < .10$ . The nature of this interaction, as shown in Table 2, was such that there was essentially no difference in RTs to words, as opposed to nonwords, when the word preceding the target-bearing word began with the dissimilar phoneme /s/. Conversely, there was a substantial difference when the item preceding the target-bearing word began with the similar phoneme /p/. A closer examination of this interaction was made via planned comparisons on the data obtained when /p/ preceded the target and, separately, when /s/ preceded the target. As expected, the analyses of variance revealed that there was no significant effect of lexical status (words, 523 milliseconds; nonwords, 535 milliseconds) when the preceding phoneme was /s/;  $F/1(1, 60) = 2.14, p < .25$  and  $F/2(1, 28) < 1$ . However, when the preceding phoneme was /p/, words were responded to significantly faster ( $\bar{X} = 549$  milliseconds) than nonwords ( $\bar{X} = 592$  milliseconds),  $\min F'(1,80) = 7.94, p < .005$ .

## Discussion

The results of Experiment II were as expected. When the word preceding the target-bearing item began with a phoneme that was similar to the target, a word superiority effect was observed. No such effect was obtained when the preceding word began with a dissimilar phoneme. This, then, is consistent with the predictions derived from the dual code model. However, for two reasons we considered it necessary to conduct an extended replication of Experiment II. First, we now had some evidence in favor of the model and some inconsistent with it; and, what's worse, both the corroborating and the disconfirming evidence were found in studies that were based on the same line of reasoning. Second, the data from Experiment I were consistent with the model only when the target phoneme was /b/ and not when it was /d/, a paradoxical result. Experiment II was also consistent with the dual code model, but it employed only /b/ targets. Perhaps the results are for some reason dependent upon the identity of the target itself. Accordingly, we conducted a follow-up study that varied the identity of the target phonemes.

## Experiment III: Phonetic Similarity with Multiple Targets

This experiment replicated the prior one in its basic logic. The target phoneme began either a word or a nonword, and the word immediately preceding the target began with a phoneme

that was either similar to the target or not. In addition, however, two target phonemes were used here, /b/ and /d/, just as in Experiment I. Thus, from the point of view of the subject, the present study was like the first experiment in that two targets were monitored for on each trial. When the target was /b/, the preceding word began with either a /p/ or an /s/; when the target was /d/, the preceding word began with either a /t/ or an /s/. (Of course, these consistencies were not pointed out to the subjects and it is unlikely that they were ever noticed given all the filler items.)

According to the dual code model, we should observe results in this study that are parallel to those in Experiment II. Subjects should be slower to respond to targets on nonwords than to those on words when the target is preceded by a word beginning with a similar phoneme. No such difference should show up when the preceding word begins with a phoneme that is dissimilar to the target. And, of course, this effect should hold across both target phonemes /b/ and /d/.

## Method

**Design and materials**—Thirty-two basic experimental sentences were constructed. In half of these sentences the target phoneme was /b/, in the remaining half the target was /d/. For each of the two target phonemes the sentence contained either a real word or a nonsense word. In addition, preceding each target-bearing item there was an adjective that began with either a highly similar phoneme (/p/ for /b/ targets; /t/ for /d/ targets) or a dissimilar phoneme (/s/ for both targets). Thus, for each target phoneme each sentence had four versions: The target-bearing word was either a real word or a nonsense word; crossed with this variable, the word immediately preceding the target-bearing item began with either a highly similar or a dissimilar phoneme. An example sentence with the target phoneme /b/ is The vacationers went to a private/secret beach/bedge to try out their new scuba equipment. An example sentence with the target phoneme /d/ is The various groups of scientists constantly argued over a technical/serious difference/deblus in their professional theories.

The preceding adjectives were equated, across /p/-initial and /s/-initial pairs, and across /t/-initial and /s/-initial pairs for frequency and syllable length. The mean Ku era and Francis (1967) frequency of the /p/-initial adjectives was 68.57, while the mean for their /s/-initial mates was 75.25. The mean frequency of the /t/-initial adjectives was 76.50, while the mean for their /s/-initial controls was 80.36. The mean length in syllables of both the /p/-initial and /t/-initial adjectives was 1.88, and it was 1.81 for both sets of the /s/-initial adjectives. The preceding adjectives were comparable in meaning (e.g., painful/stinging; tired/sleepy) and were selected so that the overall sentential meaning would be highly similar between the two sentences.

In order that each basic sentence could occur in each condition across the experiment, four material sets were constructed. Each material set contained all 32 basic sentences; 8 sentences in each material set came from each of the four conditions. In addition, 32 filler sentences were constructed. Sixteen of them did not contain a target phoneme; 8 of these contained a nonsense word and 8 did not. Another 16 filler sentences contained one of the two targets, half beginning real words and half beginning nonsense words. Balanced across the two target phonemes and the word/non-word conditions, half of the filler targets

occurred relatively early in the sentences and half occurred relatively late. All subjects were tested with a paraphrase task. The experiment was a 2 (target phoneme: /b/ vs /d/) by 2 (word type: real/nonsense) by 2 (preceding word-initial phoneme type: similar/dissimilar) by 4 (material sets) design, with the first three variables within subjects and the last between subjects. However, the particular instantiations of word type and similarity were nested within the target phoneme type (i.e., they were different for /b/ and /d/ targets, as described above).

**Subjects**—The subjects were 60 undergraduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Fifteen subjects were randomly assigned to each of the four experimental tapes (material sets).

**Procedure**—The procedure used in Experiment III was identical to that used in Experiment I.

## Results

The mean RTs from Experiment III are displayed in Table 3. Analyses of variance performed on these data, by subject and by item, revealed a significant main effect of phoneme (/b/ vs /d/) indicating that, just as in Experiment I, overall reaction times were significantly longer to the phoneme target /d/ ( $\bar{X} = 596$  millisecond) than to the phoneme target /b/ ( $\bar{X} = 543$  millisecond),  $\min F'(1,34) = 10.44, p < .005$ . In addition, there was a significant main effect due to the identity of the preceding critical phoneme. RT's were substantially longer when the target bearing word followed a /p/- or /t/-initial word ( $\bar{X} = 593$  milliseconds) than when the target followed an /s/-initial word ( $\bar{X} = 542$  milliseconds);  $\min F'(1,37) = 10.70, p < .005$ . There was a significant interaction of phoneme target and lexical status, such that /b/ words ( $\bar{X} = 514$  milliseconds) were responded to faster than /b/ nonwords ( $\bar{X} = 564$  milliseconds), yet /d/ words ( $\bar{X} = 627$  milliseconds) were responded to slower than /d/ nonwords ( $\bar{X} = 565$  milliseconds;  $\min F'(1,33) = 15.27, p < .001$ ). As can be seen in Table 3, this interaction was further exaggerated when similar phonemes occurred prior to the target phoneme, creating a significant three-way interaction between target phoneme, lexical status, and preceding phoneme,  $\min F'(1,37) = 6.08, p < .05$ . Additional planned comparisons were conducted in order to investigate the effects of preceding phoneme on the /b/ target-bearing items alone, as well as on the /d/ target-bearing items. Both analyses of variance revealed significant main effects of preceding phoneme and lexical status, as well as significant interaction terms of those two variables (all  $p$ 's  $< .05$ ).

## Discussion

Recall that in Experiment I we observed a word superiority effect when the target phoneme was /b/, and a nonword superiority effect when the target was /d/. In that study the targets were generally preceded by words that began with phonemes that were not similar to the targets. In the present experiment we replicated that result. When the initial phoneme on the word preceding the target was not similar to it, we got a word superiority effect with /b/ targets and a nonword superiority effect with /d/ targets. When the word preceding the target

began with a phoneme similar to it, these results were simply amplified. The superiority of words over nonwords was even greater when the target was /b/, and the superiority of nonwords over words was even greater when the target was /d/. If the results of Experiment I were paradoxical when the target was /d/, then the results of Experiment III might be called super paradoxical.

According to the dual code model, phonetic similarity between the target phoneme and the initial phoneme of the preceding word should have led the subjects to respond often to the postlexical code. This means that a word superiority effect should have been observed for both /b/ and /d/ targets in the similarity condition. Clearly, then, the predictions from the dual code model are not being sustained. The apparently supportive results observed in Experiment II are not general.

## Converging Operations and the Theoretical Issues

To this point we have presented the results from two lines of research testing the dual code hypothesis. This work was originally meant to test certain assumptions of the model via use of converging operations. (And, of course, we anticipated that the results would support those assumptions.) Two distinct ways of varying processing load were manipulated with the expectation that subjects would respond to the postlexical code whenever processing load was great. As it has turned out, however, neither manipulation has yielded data of the expected sort, and both have yielded results that we must view as paradoxical: a nonword superiority effect when the target was /d/.

To help clarify our present situation, recall that we began with a number of initial theoretical concerns. One important issue is the extent to which listeners compute a prelexical representation of the input signal as they process speech. A second issue is the extent to which syntactic and semantic context affects the processing of the “low level” speech signal. And a third issue has to do with the manner in which subjects carry out the phoneme monitoring task. With respect to the first question, a number of different answers have been suggested in the literature. One common answer is that phonetic segments are not a part of the code that is computed prior to lexical access (Klatt, 1980; Morton & Long, 1976; Rubín et al. 1976; Warren, 1976); from this perspective, phoneme detection is strictly a top-down process. Another view is the one we have been assuming is correct, namely that phonetic segments are computed en route to lexical access (i.e., a bottom-up view), but that subjects only respond to such units under particular sets of circumstances—the dual code hypothesis. As mentioned earlier, some of the strongest evidence in favor of the view that phoneme identification is top-down comes from the Morton and Long experiment and from the work of Blank. Another important source of evidence in favor of the top-down view comes from the work of Rubín et al. However, the data gathered in Experiments I and III both confirms and contests the latter's findings. That is, with /b/-initial targets we found evidence for the top-down view; but with /d/-initial targets the data strongly contradicted the top-down hypothesis.

To account for the contrasting data gathered in Experiments I and III above, we will adopt a new working hypothesis, present some tests of it, and critically examine the Morton and Long and the Rubin et al. experiments in light of it.

## The Prelexical Code Revisited

Suppose it is true that subjects in phoneme monitoring experiments often identify the phoneme in a top-down fashion, and that the sentence context can affect the speed of such top-down identification. If so, then we might expect to find a correlation between the time subjects take to respond to different target words when they occur in exactly the same sentence frames (assuming, of course, that such factors as frequency are held constant). That is, a common sentence frame should lead to common top-down processing for target words; this, in turn, should lead to similar RTs for those words.

To examine this assumption we computed a correlation between RTs for /d/-word targets and those for /b/-word targets when they occurred in the same sentence frames in Experiment I. The resulting  $r$  was a nonsignificant 0.161. At the same time, however, we computed the analogous correlation for non word targets (i.e., a correlation between RTs for /d/-nonword targets and those for /b/-nonword targets). This correlation was significant,  $r = .325, p < .05$ . Recall that the sentence frames for all four items were identical. Even so, the correlation among RTs was not significant for word targets, where the top-down processes might reasonably be expected to be most powerful, while it was significant for nonword targets.

Examination of the materials used quickly suggests why the latter correlation alone might have obtained. Examples of the quadruples used in that experiment are bowl, bap, dish, dap; Boston, Beggel, Dallas, Deggel; bishops, beppems, deacons, deppems. While the /b/ and /d/ words that occurred in a sentence frame were similar in frequency, number of syllables, and semantic class, they were dissimilar in their internal phonetic structure. The /b/ and /d/ nonwords, on the other hand, were similar in their internal structure—they rhymed. Thus, there was an inherent acoustic–phonetic similarity between the non word target items occurring in a sentence frame, a similarity that was not present in the word targets. It thus seems plausible to suggest that the low level phonetic environment is the important determiner of relative RTs.

Further support for this conjecture comes from an additional correlational analysis carried out on the data from our control study reported in the discussion of Experiment I. Recall that it was a study in which target items were presented in lists rather than in sentences. For those data a significant correlation ( $r = .391, p < .05$ ) was found between RTs to /b/ nonwords and their /d/ nonword counterparts. Again, these items rhymed with each other. In contrast, the correlation between the /b/ words and their /d/ word mates (similar in meaning but not in phonetic structure) was not significant,  $r = .124$ . The small correlation between /b/ and /d/ words would be expected from an analysis that emphasizes top-down effects on RTs; in fact, no correlation would be expected in lists. But the significant correlation between the /b/ and /d/ nonwords, and the fact that the pattern of correlations is



the same whether the targets occur in sentences or in lists, suggests strongly that it is the phonetic environment only that is contributing to the similarity in RTs.

## Paradox Lost

The correlational analyses suggest that the results found in Experiments I–III need reinterpretation since those studies did not always control explicitly the immediate acoustic–phonetic context of the targets. It might have been the case that the paradoxical nonword superiority effect in the case of the /d/ targets occurred because the contexts within which the two classes of items appeared were systematically different. Of course, if that is the case, then the word superiority effect in the case of the /b/ targets might *also* be due to acoustic–phonetic variables rather than to the lexical status of the item. This would be an important finding since it would constitute quite a blow to the top-down view of phoneme identification.

Given the above speculations and analyses, we decided to test directly whether the acoustic–phonetic structure of the syllable carrying the target affected systematically the RTs.

## Experiment IV: Response Times to /d/-Initial CVCs

In this study we investigated the RTs to a very large subset (essentially all) of /d/-initial CVCs. Of course, some of these CVCs were real words (e.g., /duk/ pronounced “duke”) while others were not (e.g., /dik/ pronounced “deek”). Our basic questions were whether the RT to a CVC would be affected by its status as word or a nonword; by the identity of its vowel, or by its final consonant; by some combination of these variables; or by any of them. If RTs are strongly affected by the phonetic context, then we must examine the contexts used in the earlier studies to see whether they can account for the results. Again, since the data for /d/ targets were apparently paradoxical, we chose to examine CVCs starting with /d/.

### Method

**Design and materials**—A total of 209 /d/-initial CVCs (consonant–vowel–consonant) syllables were constructed by factorially combining 11 different medial vowels with 19 different final consonants. The 11 medial vowels used were /i, I, e, ε, ae, ɔ, o, ai, u, ^, ʌ, ɜ /; the nineteen final consonants used were /p, b, t, d, k, g, m, n, ŋ, f, v, θ, s, z, ʃ, r, l, ʒ, ʒ̃, j̃/. Of these 209 CVCs, 84 were actual English words and 125 were nonwords. Lexicality was operationalized as any entry in Webster's *Seventh New Collegiate Dictionary*. In addition, 391 filler, non-target-bearing CVC syllables were constructed. The initial consonants of the filler syllables were /p, t, b, k, f, v, g, m, n/. The medial vowels and final consonants of the filler syllables were randomly chosen and combined from the sets used to construct the target-bearing CVC syllables. Of the 391 filler CVCs, 216 were actual English words and 175 were nonwords.

The 600 item list was divided into 20 sequences of 30 items each. Ten of the sequences contained 10 target-bearing CVCs and 20 filler CVCs; the other 10 sequences contained 11 target-bearing CVCs and 19 filler CVCs. In addition, within each sequence, half of the

target-bearing items were words as were half of the filler items; the others were nonwords. Ten sequences were arbitrarily chosen to be Block 1 with the remaining being Block 2. No target item occurred within the first five items on either block. The interstimulus interval was 2 seconds within sequences and 5 seconds between sequences.

**Subjects**—The subjects were 22 undergraduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Eleven subjects were randomly assigned to hear Block 1 of the experimental materials first, then Block 2, while the other 11 subjects first heard Block 2, then Block 1.

**Procedure**—Subjects were tested in groups of one to six, with the experimenter and subjects occupying adjoining rooms. Each subject was seated in a booth out of direct sight of the others.

Instructions outlining the subjects' task were recorded on tape and presented at the beginning of the experimental session. The subjects were told that they would hear a list of syllables, some of which might be real English words and others not, and to respond to syllable-initial /d/'s. They were instructed to ignore the lexical status of the syllable. Subjects were told to lightly rest the index finger of their preferred hand on the response button in front of them. Upon hearing a /d/ sound beginning any syllable in the list, they were to press the response button. They were informed that the experimenter was interested in both their speed and accuracy at listening to the list and pressing the response button, and thus that they should respond as quickly and as accurately as possible. Subjects listened and responded to four practice sequences. After the experimenter answered questions concerning any uncertainties regarding the task, the first block of the experimental tape was played. After hearing the first block, subjects were given a 5-minute rest period and then listened to the second block.

## Results

Mean correct reaction times for each of the 209 eve syllables were obtained by collapsing over subjects. These mean RTs were submitted to a three-way analysis of variance which revealed a significant main effect of vowel identity,  $F(1,10) = 4.53, p < .05$ ; however, there was no effect of final consonant identity,  $F < 1$ , nor lexical status,  $F < 1$ . There was essentially no difference in mean RTs when the target phoneme began CVCs that were actual English words ( $\bar{X} = 417$  milliseconds;  $N = 84$ ) than when the target phoneme began eves that were nonwords ( $\bar{X} = 415$  milliseconds;  $N = 125$ ).

To further examine the nature of the significant effect of medial vowel, a mean RT was obtained for each of the 11 experimental vowels by collapsing across final consonants and lexical status. These 11 means, as shown in Table 4, were then rank ordered. Visual inspection of this array suggested that reaction time increased monotonically with vowel duration. Peterson and Lehiste (1960) isolated, analyzed, and measured the intrinsic durations of all English vowel nuclei when produced in CVC syllables. A Pearson product-moment correlation between the mean vowel RTs from the present experiment and the vowel durations reported by Peterson and Lehiste (1960) revealed a significant relationship

between the two measures,  $r = .627, p < .05$ , such that RTs increased as vowel duration increased.

## Discussion

The results of Experiment IV did not give comfort to the dual code hypothesis. There was no effect on RTs due to the status of an item as a word or a nonword. One might argue, following Foss and Blank, that the failure to find such a difference was because the task was too easy for the subjects and thus did not lead them to respond at the postlexical level. However, such an argument would be extraordinarily strained given the failure to find any effects of processing load (i.e., multiple targets or phonetic similarity) on the code to which subjects responded in the earlier experiments of this series.

The present results are not random, however. It appears strongly that the RTs are affected by the immediate phonetic context within which the target appears. In particular, the structure of the syllable, especially its vowel duration, can predict RTs to its initial segment. Recall that the correlation between RTs and duration of vowels as measured by Peterson and Lehiste (1960) was 0.627, a substantial correlation considering the fact that we are dealing here with different tokens of different types spoken by a different speaker.

This finding is perhaps not so surprising when we consider a fundamental property of the speech code widely supported by research in the neighboring field of acoustic phonetics. The speech signal is often transmitted in parallel such that information about more than one phonetic segment is broadcast by the speaker at a given moment in time. Thus, to take the well-known analogy, the phonetic units within the stream of speech are not like beads on a string that can be easily dissected or even severed. Indeed, Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) have shown that cutting progressively into a syllable such as /di/, from the right-hand end, only yields sounds that are identified as a consonant plus vowel or as a non-speechlike sound, e.g., a chirp. At no point can one splice out a sound that would be identified as /d/ alone. Thus, because much of the acoustically rich information necessary for identification of a stop consonant is carried by the succeeding segment, it is reasonable that identification of the former would be affected by some characteristics of the latter. From the present data, we suspect that the length of the following vowel is one such characteristic.

In a recent paper, Sequi, Frauenfelder, and Mehler (1982) report findings that are quite similar to those found in Experiment IV and reach conclusions that are quite similar to those just suggested. These investigators varied the lexical status of target-bearing items (words vs nonwords) and asked subjects to monitor for either the initial phoneme or the initial CV syllable. Items were presented in lists. They found no effect due to an item's lexical status; but they did find a significant correlation between the time required to respond to phoneme targets and that required to respond to the syllable targets. Thus, their two findings—lack of an effect due to lexical status, and a positive correlation between RTs to initial phonemes and to their syllables—track closely the findings from Experiment IV. They conclude that “the recognition of the first phoneme or syllable in a word does not require prior lexical access to that word”; and that “the recognition of a phoneme seems to depend on the previous recognition of the syllable to which it belongs” (1982, p. 475). Whether phoneme recognition generally requires recognition of the syllable or of its vowel is not clear, but the

point is certainly well taken, given the encoded nature of the speech signal, that stop consonants at least must be identified in concert with their supporting vowel.

Recall that Rubin et al. (1976) found in a list experiment that subjects responded faster to target phonemes when they began words than when they began nonwords. It now seems quite plausible to argue that Rubin et al. observed their results because the sample of words and nonwords that they used was small and may have confounded phonetic factors with an item's status as a word or a nonword. Experiment IV examined nearly by exhaustion all of the /d/-initial CVC syllables. Thus, the results from that "sample" and from the work by Sequi et al. suggest strongly that there is no effect due to the lexical status of a target-bearing word.

Moreover, the finding that vowel duration, as measured by Peterson and Lehiste, reliably predicted the reaction times observed in Experiment IV necessitated a reexamination of the materials used in the earlier experiments here. As previously suggested, it is conceivable that the materials used in these experiments might also have confounded vowel length with the target-bearing item's status as a word or nonword. Post hoc analyses conducted upon our materials confirmed this hypothesis. First, in all three experiments vowel duration, as measured by Peterson and Lehiste (1962), correlated highly with mean RT. The obtained correlation coefficients were 0.393, 0.273, and 0.436 for Experiments I through III, respectively, all  $p$ 's < .05. (The analogous correlation was 0.440 in the list experiment described in the discussion section after Experiment I.) Second, in all these experiments, the temporal duration of the modal vowel used in the /b/-initial words was relatively *shorter* than the modal vowel used in the /b/-initial nonwords. Conversely, in Experiments I and III the temporal length of the modal vowel used in the /d/-initial words was relatively *longer* than the modal vowel used in the /d/-initial nonwords. As such, these data as well as those of Sequi et al. cast serious doubt upon the validity of the Previously observed advantage in RTs to /bi/-initial words as opposed to nonwords, while simultaneously providing a simple and logical explanation for the previously observed paradoxical /d/-word disadvantage.

Consonant with the findings of Experiment IV and those of Sequi et al. are data obtained by Swinney and Prather (1980). They found that RTs to /b/-initial CVCs were directly related to the uncertainty of the following vowel. When subjects knew what the vowel would be, they were faster to respond in the monitoring task than when the vowel was drawn from a set of four; the latter condition led to faster RTs than when the vowel was drawn from a set of eight. When subjects were given the entire syllable as a target RTs were not affected. (Also, no effect was observed due to vowel uncertainty when the target was /s/. Swinney and Prather suggest that this might be due to the difficulty of discriminating initial /s/ targets from background noise.) Swinney and Prather concluded that "at least with phoneme /b/, it appears that the detection/identification process involved in phoneme monitoring is affected by knowledge of the identity of the vowel accompanying the target phoneme" (p. 107).

It is also instructive to examine the pattern of RTs in Table 3 in the light of these arguments. Overall, we see that RTs to /d/ targets are approximately 60 milliseconds longer than those to /b/ targets, consistent with the data from Experiment I. This supports the conjecture that uncertainty regarding the word boundary slows RTs to /d/ targets. In addition, we can now

interpret the significant three-way interaction between target phoneme, lexical status, and preceding phoneme: For each target phoneme the fastest RT occurred with the dissimilar preceding phoneme /s/ and when the vowel following the target was short. This happened with word targets for /b/'s and nonword targets for /d/'s. We also see that there is an increment in RT from that "base" due to both similarity of preceding phoneme and to vowel duration. However, the combination of these two yields an even bigger effect than one would expect from a simple additive model, for example,  $RT = f(\text{vowel, preceding target, vowel} \times \text{preceding target})$ . This effect could be accounted for by a model in which (a) a similar preceding phoneme led subjects to raise their thresholds for responding to the target (i.e., the near occurrence of a false alarm leads subjects to require more information consistent with the target before they respond), and (b) the reasonable assumption is made that the growth rate of such information across time is faster when the target-bearing syllable contains a short vowel.

To summarize, these experiments suggest that subjects recognize the CV portion of the syllable before responding, and that the time required to make such a response is determined by the uncertainty of the vowel as well as by the temporal factors within the syllable that determine vowel recognition.

### Semantic Context Effects Reexamined

Earlier we noted that some of the strongest evidence in favor of a top-down interpretation of phoneme identification came from the work of Rubin et al. and from the work of Morton and Long. The results from Experiment IV suggest strongly that the Rubin et al. evidence for top-down processing is quite weak. Our data show that RTs are not affected by the lexical status of a word, but are affected by its phonetic structure. Given this finding, let us examine critically the data from Morton and Long and its various replications. Recall that Morton and Long found that RTs to target phonemes were significantly faster when the target-bearing word was predictable given the sentential context. However, an inspection of their materials suggests that semantic predictability may have been confounded (negatively correlated) with the duration of the vowel in the target words, as well as with the existence of initial consonant clusters in the target words. Such clusters would mean a delay in the time for the vowel (or syllable) to be identified. The majority of target-bearing words used by Morton and Long in highly predictable semantic contexts is composed of medial vowels which have been classified by Peterson and Lehiste (1960) to have relatively short temporal durations. On the other hand, a majority of the low predictability target-bearing words was composed of medial vowels classified as having relatively long temporal durations. Furthermore, within the set of high predictability target-bearing items only four target phonemes occurred in initial consonant clusters, whereas there were twice as many for the set of low predictability target-bearing words. Since the unpredictable words were also those with consonant clusters and relatively long vowels, we do not know whether the longer RTs for targets on those items were due to their low predictability or to their phonetic structure.

## Experiment V: Response Times to “Predictable” and “Unpredictable”

### Words Presented in Lists

Experiment V was a control study in which we presented the target-bearing words from the Morton and Long experiment. However, the target words were presented in lists rather than in sentences. The logic of this control experiment is straightforward. If the RT differences in the Morton and Long study were due to the predictability of the target-bearing word, then we should only observe a difference in RTs when those words occur in their appropriate contexts; for it is only in those cases that such predictability can affect phoneme identification via the top-down route. When the target words occur in isolation, as in the present study, the probability of top-down identification due to predictability is reduced to zero. Therefore, one would not expect to see any difference between the “predictable” and the “unpredictable” items when they occur in lists. On the other hand, if the earlier results were due to the phonetic structure of the target-bearing words themselves, and not to the sentential context, then we might expect to see a difference in RTs to “predictable” and “unpredictable” items even when the words occur in isolation. Accordingly, we conducted an experiment in which the subjects were presented with the target words from the Morton and Long study. These words were presented in lists and subjects were asked to monitor for the target phoneme beginning these words.

### Method

**Design and materials**—This study used 20 of the target-bearing words which were taken from Morton and Long's (1976) experimental sentences; these were also used by Foss and Blank (1980; Experiment IV). Half of the target-bearing words originally appeared in those studies in contexts such that they were highly predictable, while half had appeared in contexts such that they were not predictable. For example, the word *beer* is highly predictable, and the word *brandy* is not predictable, in the context, “He had a drink of beer/brandy in the hope that it would cheer him up.” The present experiment did not present the target-bearing words in sentences and thus did not manipulate transitional probability as had Morton and Long and the replication conducted by Foss and Blank. However, we were interested in the RTs to the two sets of words that had been labeled as high and low predictable in those studies. For convenience, we will continue to refer to them as high and low predictable items, though it should be understood clearly that no context, and hence no difference in predictability, was manipulated here.

The words making up the two groups of items were equated for word frequency. The mean frequency per million items for the high predictability nouns was 81.6, while for the low predictability nouns it was 81.0 (Thorndyke & Lorge, 1944). The mean length in syllables of the high predictability nouns was 1.3, and of the low predictability nouns 1.4. Five target phonemes were used (/b,d,t,p,k/) with four high predictability and four low predictability nouns starting with each of these phonemes. Blocks containing six words each were formed. For each target phoneme, 22 blocks were constructed: 4 practice blocks containing one target-bearing nonexperimental word and 5 filler words; 8 experimental blocks containing one target-bearing experimental word and 5 filler words; 8 filler blocks containing one target-bearing nonexperimental word that was a non-noun, and 5 filler words; 4 catch blocks

containing no target bearing words but rather one word with a highly similar initial phoneme (e.g., /p/ for /b/ target trials) and 5 filler words. To approximate the composition of the actual experimental words, all filler words were chosen such that their mean frequency was approximately 75, mean length was 1.5 syllables, and 40% were nouns.

The experimental words occurred in positions 3 or 4 of their respective blocks with the constraint that a high predictability noun and its low predictability mate (as used in Morton and Long's stimuli) occurred in the same ordinal position in their respective blocks. The target-bearing nonexperimental words occurred in positions 1, 2, 5, or 6 of their respective blocks. Filler words were randomly arranged around both the experimental and nonexperimental target-bearing words. However, we felt that it was also necessary to scrutinize further the final ordering and to rearrange any target bearing word that had been assigned to a block containing filler words that were even slightly semantically related. This resulted in rearranging two nonexperimental words. The interstimulus interval was 1, 2, or 3 seconds randomly occurring between each item in the list, with the constraint that a 2-second interstimulus interval always occurred before an experimental word. For each target phoneme, blocks of each type were randomly arranged into one sequence with the constraints that the four practice blocks always occurred first and that no more than two of the same type of block (e.g., experimental, filler, or catch) occurred successively.

All five sequences (110 blocks) were recorded by a female speaker on one channel of a tape. A pulse, inaudible to subjects, was placed on the second channel of the tape at the beginning of the target phoneme. The pulse started a timer which was stopped when subjects pressed a button.

**Subjects**—The subjects were 18 under-graduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement.

**Procedure**—Subjects were tested in groups of one to six, with the experimenter and subjects occupying adjoining rooms. Each subject was seated in a booth out of direct sight of the others.

Instructions outlining the subjects' task were recorded at the beginning of the experimental tape which was presented binaurally over headphones. Subjects were told to lightly rest the index finger of their preferred hand on the response button in front of them. They were told that they would hear a list of words and that they were to listen for a word that began with the specified target phoneme. Upon hearing the initial phoneme target, they were to press the response button as quickly as possible. Subjects were told that the words in the lists were not chosen or arranged in any particular way and to try to listen to the individual words without associating any word with the word(s) that preceded it in the list. Before each phoneme sequence began, subjects were informed of the particular target phoneme (e.g., "Now listen for the sound of /b/ as in Bob or Boston or Brazil that begins any word in the list.") After the experimenter answered questions clarifying any uncertainties regarding the instructions, the experimental tape was presented.

## Results

A one-way analysis of variance performed on the mean correct reaction times from this experiment revealed a main effect of “semantic predictability,”  $F(1, 17) = 33.82, p < .001$ . Phoneme targets beginning words that had been presented by Morton and Long in high predictability contexts were responded to much more rapidly ( $\bar{X} = 508$ ) than phoneme targets beginning words that had been presented in low predictability contexts ( $\bar{X} = 560$ ), a mean difference in reaction times of 52 milliseconds.

Foss and Blank (1980), using the same subset of items from the Morton and Long stimulus set, but with the items presented in sentences, reported a mean difference of 47 milliseconds with the same pattern of results, that is, phoneme targets, beginning high predictability words were responded to faster. A post hoc analysis on the present data revealed that phoneme targets occurring in consonant clusters were responded to an average of 106 milliseconds slower than those not occurring in clusters. Although a similar post hoc analysis performed upon the RTs to target phonemes occurring in the different vowel contexts would be interesting, we do not feel that an average of three mean data points per vowel would support the validity of such an analysis.

## Discussion

The results from Experiment V were quite clear. RTs to the “predictable” items were significantly faster than those to the “unpredictable” ones, even though predictability did not vary between the items. The magnitude of the effect was of the same order as that observed when the target items occurred in context. The only reasonable conclusion appears to be that the target-bearing words differ in some intrinsic, low level fashion in such a way as to yield the RT difference. And, as noted above, the words do in fact differ in their phonetic structure. The “predictable” ones generally have shorter vowels and fewer initial consonant clusters than do the “unpredictable” ones. It thus seems clear that the RTs differences observed here are due to the phonetic structure of the target items. By parity of argument, the results observed by Morton and Long and by Foss and Blank were probably also due to the phonetic variables and not, as they thought, to the predictability of the target items. These studies cannot be used to argue for predictability effects.

There is one important study which is not subject to the criticism that phonetic variables are confounded with semantic predictability, namely the work of Blank (Note 1, Experiment 1). In her experiment the target-bearing word was held constant and its predictability was varied by manipulating the prior sentence context. Blank found faster RTs for predictable targets than for unpredictable ones, a result consistent with a version of the dual code hypothesis. It is possible, though, that her results may have been affected by overall differences in the ease of sentence processing or by the fact that subjects were listening on some trials for target phonemes that had actually been excised from the signal and replaced by noise. Clearly, given Blank's results, additional work would be required before we could claim that there is no effect due to semantic predictability of the target-bearing word. Such work is presently underway.



## Summary: On the Role of Context in the Perception of Speech

We began this paper with a number of questions. One was whether listeners can be shown to compute a low level representation of the speech code en route to lexical access. A second, and related, question was whether the computation of such a low level code (assuming one *is* developed during comprehension) is affected by the ongoing syntactic and semantic analyses of the input. A third question was whether the dual code model of phoneme identification could be sustained. Let us review our results and address each of these issues briefly.

Five experiments have been presented here. We began by testing in various ways predictions from the dual code model of phoneme identification. According to this model listeners respond to target phonemes at both the prelexical and postlexical levels depending upon the circumstances. Subjects are more likely to respond to the postlexical code when their processing resources are taxed, according to the model. In the first experiment we taxed the subjects' memories by asking them to monitor for multiple targets. Contrary to expectations, we did not find evidence in favor of the model. Indeed, we found some anomalous data; namely, we observed evidence for an occasional nonword superiority effect. In the second and third studies we manipulated the subjects' processing load by immediately preceding the target-bearing words with others that began with phonemes which were either phonetically similar or dissimilar to the actual targets. Again we found no evidence that subjects respond to a postlexical representation of the target phoneme when their processing resources are taxed. And again we found some evidence in favor of a paradoxical nonword superiority effect. In Experiment IV we tested subjects' RTs to nearly all /d/-initial CVCs in English. No difference in RTs was observed as a function of the item's status as a word or nonword. Moreover, post hoc analyses showed that RTs were affected by the fine phonetic structure of the target-bearing item; in particular, RTs were positively correlated with vowel duration. This result led us to speculate that the earlier data—including the paradoxical nonword superiority effects—could be accounted for by the low level phonetic structure of the target-bearing items. Correlational analyses sustained that conjecture. We also speculated that the important data gathered by Morton and Long could be accounted for in the same fashion. In Experiment V we found that the words used by Morton and Long in “predictable” contexts were responded to faster even when they occurred out of context.

To return to the theoretical issues, all of the evidence obtained in this series of experiments is consistent with the view that listeners do compute a prelexical representation of the speech signal. Both here and in the data gathered by Sequi et al., RTs in the phoneme monitoring task were not affected by the lexical status of the target-bearing word. (Also, see related work testing phoneme monitoring in noise described in Foss, Harwood, & Blank, 1980; their data are probably best interpreted as being consistent with this view.) These results suggest that listeners are able to decide whether an item matches the description of the target before the target-bearing word itself has been identified. In other words, responses are initiated on the basis of a prelexical representation of the stimulus. It must be noted, however, that the exact form of the representation cannot be deduced from the present data, multiple alternatives are consistent with it. For example, such a representation might be in terms of CV units or in terms of phonetic segments. In the case of the former units, however,

the subjects in our experiments must have the ability to decompose the CV unit rapidly into something quite like phonetic segments in order to carry out the monitoring task.

The data from the present research speak directly, if not definitively, to the question of whether syntactic and semantic analyses affect directly the processing of the low level phonetic (or CV) code. Some theorists have proposed that there is direct influence of this sort. For example, Morton and Long (1976), p. 44) write that “higher-order processes provide contextual cues which interact with sensory information,” such that “the more contextual cues are available the less sensory information is required.” Impressive data in support of that conjecture were gathered by Morton and Long in the paper cited. We have shown, however, that their results may have been due to the low level phonetic context within which the targets occurred, and not to the higher-order processes they credited. Of course, such an interaction might exist, we certainly have not proved that it does not; to date, though, the interaction has not been unequivocally demonstrated in studies of sentence comprehension. Work is under way to examine this matter directly.

Acoustic phoneticians and workers in the area of speech perception have known for many years that stop consonants lack invariant cues for recognition (e.g., Delattre, Liberman, & Cooper, 1955). The physical representation of such a consonant is contingent upon its neighbors, though perceptually the same consonant is reliably identified. Work using phoneme monitoring suggests that identification times reflect the extent to which the speech signal is encoded, and thus can be used to investigate which neighbors (e.g., following vowel) affect recognition.

Finally, the results of this series of experiments do not yield unalloyed support for the dual code model of phoneme monitoring. Foss and Blank proposed such a model largely to be able to account for the (apparent) sentential context effects on RTs demonstrated by Morton and Long, and for the (apparent) effects of lexical status of the target-bearing word on RTs presented by Rubin et al. We have seen that an alternative account of these two sets of results is more plausible. The only data remaining in support of the model are those gathered by Blank (Note 1). The evidence that listeners use a postlexical code when responding in the monitoring task has thus shrunk considerably.

It seems a fair summary of the present research to say that the evidence favors the view that subjects respond on the basis of a low level, prelexical code when they carry out the phoneme monitoring task. Though a dual code model could still be defended, the present evidence supports a more unitary picture of the monitoring process. This is useful information to have since it may both simplify the interpretation of existing studies that have made use of the technique and aid in the design of future studies since theorists will be able to state with greater confidence what variables affect phoneme monitoring RTs. Though further work remains, we believe that the present experiments are a whetstone on which this experimental tool has been sharpened.

## Acknowledgments

This work was supported in part by Grants MH29891 and MH15744 from the National Institute of Mental Health, by Grant BNS-03889 from the National Science Foundation, and by a Sloan Foundation grant to the Center for

Cognitive Science at the University of Texas. A report of Experiment IV was presented at the meeting of the Southwestern Psychological Association, Dallas, Texas, 1982. The authors thank Randy Diehl for his very helpful advice and insights, and Jean Newman-Charlton and an anonymous reviewer for helpful comments on an earlier version of this paper.

## References

- Delatire PC, Liberman AM, Cooper FS. Acoustic loci and transitional cues for consonants. *Journal of the Acoustic Society of America*. 1955; 27:769–773.
- Dell GS, Newman JE. Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*. 1980; 19:608–623.
- 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*. Erlbaum; Hillsdale, N.J.: 1980.
- Foss DJ, Swinney D. On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:246–257.
- Healy A, Cutting J. Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*. 1976; 15:73–83.
- Klatt, D. A new look at the problem of lexical access. In: Cole, RA., editor. *Perception and production of fluent speech*. Erlbaum; Hillsdale, N.J.: 1980.
- Ku era, H.; Francis, WN. *Computational analysis of present-day English*. Brown Univ. Press; Providence, R. I.: 1967.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- McNeill D, Lindig K. The perceptual reality of phonemes: Syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:431–461.
- Morton J, Long J. Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior*. 1976; 15:43–51.
- Peterson GE, Lehiste I. Duration of syllable nuclei in English. *The Journal of the Acoustical Society*. 1960; 32:693–703.
- Rubin P, Turvey MT, Van Gelder P. Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics*. 1976; 19:394–398.
- Savin HB, Bever TG. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*. 1970; 9:295–302.
- Sequi J, Frauenfelder U, Mehler J. Phoneme monitoring and lexical access. *British Journal of Psychology*. 1981; 72:471–477.
- Swinney DA, Prather P. Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception and Psychophysics*. 1980; 27:104–110. [PubMed: 7375332]
- Thorndyke, EL.; Lorge, I. *The teacher's word book of 30,000 words*. Bureau of Publications, Teacher's College, Columbia University; New York: 1944.
- Warren RM. Perceptual restoration of missing speech sounds. *Science*. 1970; 167:393–395. [PubMed: 4311809]
- Warren, RM. Auditory illusions and perceptual cesses. In: Lass, NJ., editor. *Contemporary issues in experimental phonetics*. Academic Press; New York: 1976.
- Winer, BJ. *Statistical principles in experimental design*. McGraw–Hill; New York: 1962.

## Reference Notes

- Blank, MA. Unpublished doctoral dissertation. University of Texas at Austin; 1979. Dual-mode processing of phonemes in fluent speech.

**Table 1**

Mean Reaction Times (msec) from Experiment I

Target-bearing item	Target phoneme	
	/b/	/d/
Word	471	569
Nonword	508	523

**Table 2**

Mean Reaction Times (msec) from Experiment II

Target-bearing item	Preceding phoneme	
	/p/	/s/
Word	549	523
NonWord	592	535

**Table 3**

Mean Reaction Times (msec) from Experiment III

Target-bearing item	<u>Target phoneme /b/: Preceding phoneme</u>		<u>Target phoneme /d/: Preceding phoneme</u>	
	/p/	/s/	/t/	/l/
Word	529	499	659	594
Nonword	614	515	570	560

**Table 4**

Mean Reaction Times to /d/ Initial CVC Syllables Grouped by Identity of Medial Vowel

Vowel	Example CVC (phonetic transcription)	Example CVC (English orthography)	Mean RT (msec)
/u/	/dʊn/	dune	400
/ɛ/	/dɛn/	den	401
/ʌ/	/dʌn/	done	408
/o/	/dɒn/	doan	408
/e/	/den/	dane	414
/ɪ/	/dɪn/	din	416
/ɚ/	/dɚn/	dern	420
/i/	/din/	dean	420
/ɔ/	/dɔn/	dawn	420
/aɪ/	/daɪn/	dine	431
/ae/	/daen/	dan	434