COGNITIVE PSYCHOLOGY 12, 1–31 (1980)

Identifying the Speech Codes

DONALD J. FOSS AND MICHELLE A. BLANK
University of Texas at Austin

Models of speech processing typically assume that speech is represented by a succession of codes. In this paper we argue for the psychological validity of a prelexical (phonetic) code and for a postlexical (phonological) code. Whereas phonetic codes are computed directly from an analysis of input acoustic information, phonological codes are derived from information made available subsequent to the perception of higher order (word) units. The results of four experiments described here indicate that listeners can gain access to, or identify, entities at both of these levels. In these studies listeners were presented with sentences and were asked to respond when a particular word-initial target phoneme was detected (phoneme monitoring). In the first three experiments speed of lexical access was manipulated by varying the lexical status (word/nonword) or frequency (high/low) of a word in the critical sentences. Reaction times (RTs) to target phonemes were unaffected by these variables when the target phoneme was *on* the manipulated word. On the other hand, RTs were substantially affected when the target-bearing word was immediately *after* the manipulated word. These studies demonstrate that listeners can respond to the prelexical phonetic code. Experiment IV manipulated the transitional probability (high/low) of the target-bearing word and the comprehension test administered to subjects. The results suggest that listeners are more likely to respond to the postlexical phonological code when contextual constraints are present. The comprehension tests did not appear to affect the code to which listeners responded. A "Dual Code" hypothesis is presented to account for the reported findings. According to this hypothesis, listeners can respond to either the phonetic or the phonological code, and various factors (e.g., contextual constraints, memory load, clarity of the input speech signal) influence in predictable ways the code that will be responded to. The Dual Code hypothesis is also used to account for and integrate data gathered with other experimental tasks and to make predictions about the outcome of further studies.

This paper is concerned with the nature of the perceptual codes that are developed during the act of understanding sentences. Our primary aims are to specify some of these codes, to present a theory of the ways in which listeners' response systems can gain access to the codes, and to describe a set of experiments that bear on the theory. In addition, we will show how the theory accounts for data gathered with a variety of "on line" measures of speech processing, e.g., phoneme monitoring, shadowing, and mispronunciation detection.

The work described here was supported by Grant MH29891 from the National Institute of Mental Health. The authors wish to thank Anne Cutler, Merrill Garrett, Bjorn Lindblom, and especially David A. Harwood, each of whom has contributed a useful suggestion or criticism.

Requests for reprints should be sent to Donald J. Foss, Department of Psychology, University of Texas at Austin, Austin, TX 78712.

0010-0285/80/010001-31$05.00/0

Understanding spoken language requires that an acoustic signal be transformed into a coded representation within the semantic system of the listener. This process is typically described in terms of stages, each stage transforming the signal from one coded representation into another. The codes most often postulated are those at the acoustic, phonetic, phonological, syllabic, and lexical levels (see Studdert-Kennedy, 1974). Empirical work has been aimed at examining the validity of these codes and the processing mechanisms that supposedly compute them. For example, Wood (1975) has presented both behavioral and EEG data supporting the distinction between the acoustic and the phonetic codes during speech perception. Similarly, Studdert-Kennedy, Shankweiler, and Pisoni (1972) argued for the existence of these two types of codes on the basis of dichotic listening data. Additional evidence can be found to support the other hypothesized codes (e.g., Massaro, 1974; Warren, 1976). However, there have been a variety of answers given to questions about the nature of the perceptual codes and the time course of their development during comprehension. For example, theorists have differed concerning whether or not phonetic segments are actually computed by the speech perception mechanisms. Recently Klatt (1980) has argued that such segments are not psychologically realized during speech perception. Similarly, Warren (1976, p. 409) has concluded that, "While phonemes are constructs useful for transcribing and analyzing, they are without direct perceptual basis . . . phonemes seem to have no direct relevance to perceptual processes leading to the comprehension of speech." The work to be reported here speaks to this controversy. One of our aims is to sharpen the relevant questions and to provide new data that bear upon them.

For our present purposes, questions about the perceptual status of phonetic segments are pivotal. We can define two broad classes of perceptual models, each of which has numerous possible instantiations. 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. Warren states clearly that the correct model is a member of the latter class. We will evaluate these two classes of speech decoding models.

Integral to our primary concern about the role of phonetic segments in speech perception is a question about the way in which subjects carry out the task known as phoneme monitoring. In this task subjects are asked to listen to a sentence and to respond by pushing a button if and when a previously specified target phoneme occurs in the sentence. Reaction times (RTs) to detect the target phoneme are often correlated positively with the difficulty of processing the sentence at the point where the target phoneme occurs. Two alternatives for the way in which subjects carry out this task have been considered. They correspond roughly to the two classes of perceptual models introduced above. One alternative is that listeners can respond directly to the phonetic code that is derived by the speech perception mechanisms from the acoustic information. A second alternative is that listeners can respond to a target phoneme only subsequent to accessing the word in the mental lexicon that carries the target. According to the latter alternative, listeners cannot gain access directly to phonetic segments. Instead, subjects in a phoneme monitoring experiment decide whether a word contains the target phoneme by examining the phonological representation of the word after they have retrieved it from the mental lexicon.

The connection between models of speech decoding and the manner in which subjects carry out phoneme monitoring is a close one. If listeners do not compute phonetic segments as they recognize speech, then it follows that subjects can respond to target phonemes only subsequent to lexical access. If, however, the alternative class of perceptual models is correct, that is, if subjects do compute phonetic segments prior to lexical access, then it may be possible (though certainly not necessary) for them to respond to a target phoneme prior to accessing the word containing it. Thus, if it could be shown that listeners can respond to targets prior to lexical access, that fact alone would be very strong evidence in favor of this class of perceptual models.

In an earlier paper concerned with the point at which listeners can respond to target phonemes during sentence processing, Foss and Swinney (1973, p. 253) concluded that "the monitoring task does not tap into the comprehension processes at a level that corresponds to immediate perception." They concluded, in other words, that subjects respond to targets subsequent to lexical access. This conclusion was in contrast to an earlier assumption (e.g., Foss & Lynch, 1969) that subjects can respond to a target phoneme on the basis of either its acoustic or its phonological form. More recently, Morton and Long (1976) presented data consistent with Foss and Swinney's conclusion that phonemes are identified and responded to after lexical access. In their experiment subjects were presented with sentences such as *He sat reading a book/bill until it was time to go home for his tea,* and were asked to push a button when they heard a word-initial /b/. The word carrying the target was either predictable from the context (e.g., *book*) or not (e.g., *bill*). Morton and Long found that the times to respond to the target were significantly shorter when the word carrying the target was predictable (i.e., had a high transitional probability). In their study the target-bearing words were equated for frequency.

In order to explain the observed effects of word predictability on responses to phonemes initiating those words, it is necessary to make two assumptions: first, that highly predictable words are more rapidly accessed than are less predictable words; and second, that listeners identify the target phoneme and respond to it only after the word has been retrieved. Indeed, Morton and Long concluded that lexical access had occurred before phoneme identification took place, a conclusion consonant with that of Foss and Swinney. They stated, "It must be, then, that the identification of the target phoneme followed or competed with the word identification" (p. 48).

We are concerned with raising again the issue of where within the processing system the listener can gain access to phonological information. Determining this point may help us decide among the two classes of speech perception models we have discussed. Also, it will help us understand the workings of phoneme monitoring and other "on-line" measures of sentence processing. In general,' theorists need to understand how measurement techniques interact with the phonemena of interest if they are to interpret data correctly.

Further discussion of these issues will be aided by making a terminological distinction. When listeners comprehend an aurally presented sentence, they extract its phonological information. It is important to note that the term "phonological information" as used here is theoretically vague. As the input signal is processed by the listener there may be a number of points at which it would be reasonable to say that phonological information had been recovered. For example, there is a point at which the acoustic signal has been transformed into a code having linguistic significance. Here, some of the segmental and perhaps featural information has been assigned to the signal, but the lexical item has not yet been identified. We will call this a *phonetic representation* of the input. If, in the phoneme monitoring task, the listener responds to the phonetic representation, we will say that the target phoneme has been phonetically identified. The phonetic representation is, we conjecture, the (partial) basis for lexical access. Each word in the mental lexicon has associated with it information of varying sorts: syntactic, semantic, and, of special interest to us here, phonological. As soon as a word has been accessed in the mental lexicon, an abstract, *phonological representation* of the word becomes available to the listener. This, then, is a second point at which we could say that phonological information has been recovered. If the listener engaged in phoneme monitoring responds on the basis of this representation, we will say that the target phoneme has been phonologically identified. Note that phonetic representations are derived from the acoustic signal and are not dependent upon lexical retrieval, while phonological representations are typically the direct result of such retrieval. According to both Foss and Swinney and Morton and Long, phoneme monitoring occurs to the phonological representation of the input.

The results of two studies conducted earlier in our laboratories have led us to conjecture that subjects sometimes respond on the basis of the phonetic representation of the input. In those studies listeners apparently were able to respond before the target-bearing word was retrieved from the mental lexicon. Accordingly, we will adopt as a working hypothesis the claim that subjects *can* respond to the phonetic representation, as just defined. We will refine this claim below. The hypothesis says that subjects can respond to relatively early or low-level entities that are computed by the speech perception mechanisms during running speech. If they can identify these entities, then it makes available to us the possibility of gathering relatively direct evidence about the units of speech perception. And, as noted, the existence of such an ability has implications both for the type of perceptual model that is correct and for the inferences that can be made from results of phoneme-monitoring experiments. Experiment I was designed to test directly whether listeners can identify phonemes on the basis of phonetic information—information that they have available before they have accessed the word containing the target phoneme.

EXPERIMENT I

The logic behind Experiment I is simple and straightforward. If phonemes can be identified only on the basis of phonological representations, i.e., the code that becomes available after lexical access, then any variable that affects the time it takes to retrieve the target-bearing word should also affect phoneme-monitoring RTs. On the other hand, if phonemes can be identified on the basis of phonetic representations, i.e., those that are derived by the processing mechanisms from the acoustic signal and which can arise without lexical access, then a variable that affects retrieval time for the target-bearing should not affect response times.

In this experiment we manipulated a variable that should greatly affect lexical access time—namely, whether or not the target-bearing item was in the listener's mental lexicon at all. Subjects were aurally presented with sentences and were asked to listen for a word-initial target phoneme. Within some sentences the target-bearing "word" was not a word at all. Instead it was a nonsense word (a legal, but novel phonological sequence). When listeners encounter a nonsense word they cannot be successful in retrieving it from the mental lexicon since it is simply not there. Consequently, an abstract phonological representation of the input word does not immediately become available.[1]

[1] Such a representation may be constructed by the listener from the phonetic representation via a set of rules; but this is a different, and slower, route to the phonological representation than that which occurs via the retrieved lexical item.

Both Foss and Swinney as well as Morton and Long conjectured that phoneme identification occurs subsequent to word identification and that it depends upon the phonological information that is stored in the lexicon. If this is true, then response time to a target phoneme that is carried by a nonsense word should be slow relative to RTs for targets that begin real words. In contrast, if subjects can identify and respond to phonemes on the basis of a phonetic representation, then we would expect a smaller difference (or none at all) in RTs to phoneme targets beginning nonsense words vs those beginning real words. According to our working hypothesis, it is the latter state of affairs that will hold in the present experiment. We expect that subjects can, in fact, respond on the basis of information derived from the phonetic code; therefore, little or no difference in RTs will result when the target is carried by a nonsense word vs a real word.

The above prediction depends upon a further assumption, namely, that a phonetic representation of the input can be derived from the acoustic signal equally rapidly for both words and nonwords. This assumption is certainly dubitable. However, if the experiment comes out as expected, the assumption as well as the hypothesis will have been corroborated.

A second variable was manipulated in Experiment I. For half of the trials the target phoneme was carried by the word or its "matched" nonword (the *On* condition); for the other trials the target phoneme was carried by the next word in the sentence (the *After* condition). We have predicted that there will be little or no difference in RTs when the target is on the word vs the nonword. However, the situation is quite different when the target is after them. The basis for this difference primarily has to do with the position of the target phoneme within its carrier word. Recall that the subject is asked to respond to word-initial target phonemes; therefore, position as well as identification information is required before an accurate response can be made. How is position information determined? One way is for the subject to access the target-bearing word in the mental lexicon. Since word boundary information is inherent in the phonological representation of that word, the subject would then know that the target phoneme is word-initial. If this were the only way that position information could be determined by the listener it would lend strong support to the view that a subject could respond to a word-initial target phoneme only after accessing its carrier word. However, in principle it is also possible for a phoneme to be identified as word-initial without having to access the word that it begins. A listener may be able to determine that a phoneme is word initial if he or she knows that the immediately preceding phoneme is word final. That is, as soon as the listener has determined that a phonetic sequence constitutes a word, then a word boundary can be assigned. According to this view, word boundary assignment is in part a "top down" process. When a word boundary has been identified in this way, then the listener knows that the next phonetic segment is word initial. If that segment matches the target, the listener can respond. In this case it is not necessary to access the word initiated by the target.

In Experiment I the target phoneme sometimes occurred after a real word and sometimes after a nonword. In the latter case, the second of the above two ways of determining whether a phoneme is word initial is not available to the listener, and the first is severely slowed. The second way is not available simply because the nonsense item does not occur in the mental lexicon. Hence, listeners will be quite uncertain about where this item ends and where the next (real) target-bearing word begins. The word boundary cannot be determined in a top-down fashion under these circumstances. Without such word boundary information, the listener will not be sure whether to initiate a response even if the phonetic segment is identified as a target. Of course, the listener will be able to determine that a phoneme is word-initial when he or she gains access to the stored phonological representation of the target-bearing word. But this access will be slowed because of the prior occurrence of the nonword which causes difficulty in word boundary assignment. According to this analysis, then, we predict that RTs to respond to word-initial target phonemes will be longer when the target occurs after a nonword than when it occurs after a real word. To summarize, we expect that Experiment I will result in an interaction: no difference in RTs to target phonemes when they occur on a real word vs a nonsense word, and shorter RTs to target phonemes after real words than after nonsense words.

Method

Design and materials. Thirty-six basic experimental sentences were constructed. Each sentence had four versions: a sentence contained either a single nonsense word or only real English words; crossed with this variable, either the nonsense/real word or the word immediately following it began with the target phoneme. 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 36 basic sentences: one-fourth of the 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: nonsense word/real) × 2(target position: *on* nonsense-real word/ *after* nonsense-real word) × 4(material sets) factorial, with the first two factors being within subjects and the last being between subjects.

The nonsense words used in this experiment were derived from the real English words which they replaced in the experimental sentences. Each nonsense word shared with its counterpart the same initial phoneme, syllabic structure, and word stress pattern. A few examples of the nonsense/real word pairs are: *gatabont/government; burtle/babble; dackulous/dangerous.* The nonsense/real word pairs were counterbalanced across three word classes: noun, verb, and adjective. In addition, both the noun and adjective nonsense/ real word pairs were counterbalanced across subject and object position. An example experimental sentence with /g/ as the target phoneme on the nonsense/real word, and /p/ as the target phoneme after the nonsense/real word is:

At the end of last year, the *g*overnment/*g*atabont *p*repared a lengthy report on birth control.

Thirty-six filler sentences were constructed. Twelve fillers did not have a target phoneme; six of these contained a nonsense word and six did not. Another 12 fillers contained a nonsense word and a target phoneme; the target occurred well before the nonsense word for half of these sentences and well after it for the other half. The final 12 fillers contained only real English words; the target phoneme occurred early in six of them and late in the remaining six. The 72 sentences were randomized, with each basic sentence occurring in the same position for all material sets.

A female speaker recorded each of the four materials 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 each target phoneme. The pulse started a timer which was stopped when subjects pressed a button.

Subjects. The subjects were 32 undergraduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Eight subjects were 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 subjects' task were recorded at the beginning of each experimental tape. The instructions and the test sentences were 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 to listen for a word-initial sound (e.g., "/bə/ as in Bob") and to press the button as quickly as possible when they heard it. A trial consisted of the word "ready," specification of the target phoneme, and presentation of the test 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 task of pressing the button when they heard the target sound. Subjects were also told that the occurrence or nonoccurrence of the target sound was not determined by the presence or absence of a nonsense word in the sentence. Following the instructions, subjects were given three practice sentences. After the experimenter answered questions clarifying any uncertainties regarding the instructions, the experimental and filler sentences were presented.

Subjects were forewarned in the instructions that a comprehension test would be administered at the end of the experiment. This instruction emphasized the importance of paying close attention to the sentences. Immediately following the presentation of all the test sentences, subjects were given a printed comprehension test. The comprehension test was actually a recognition task consisting of 24 sentences. Half of these sentences were old, the subjects had heard them during the experiment, and half were new. Subjects were instructed to state whether each sentence was old or new. All of the old sentences were chosen from among the fillers which contained only real English words. These old fillers were the same in each material set. All of the new sentences also contained only real English words. Half of the new sentences were derived from actual filler sentences. They were semantically different but superficially similar in that either many of the words that were used were identical to those that occurred in the originally presented filler sentence, or the derived filler sentence was structurally similar to the original sentence. The following is an example of an actual filler sentence that subjects heard during the experiment and a new sentence that occurred on the comprehension test:

Filler: The football players found the coach extremely unfair so the team decided to strike.

Test: The young couple found the lawyer extremely unfair so they decided to fire him.

The other half of the new sentences were not related to any of the sentences presented during the experiment. Only the RT data obtained from subjects who made six or fewer errors on the 24-item comprehension test were used.

Results

The mean RTs in the four conditions of the experiment are shown in Table 1. Both of the variables led to significant main effects. The word-type variable (nonsense vs real word) had $F_1(1{,}28) = 30.88$, $p < .001$; $F_2(1{,}35) = 10.59$, $p < .003$; and $minF'(1{,}53) = 8.36$, $p < .01$. The target position variable (On vs After) had $F_1(1{,}28) = 222.37$, $p < .001$; $F_2(1{,}35) = 30.20$, $p < .001$; and $minF'(1{,}46) = 26.54$, $p < .001$. The interaction between these two variables was also reliable, $F_1(1{,}28) = 26.28$, $p < .001$; $F_2(1{,}35) = 9.87$, $p < .005$; and $minF'(1{,}58) = 6.96$, $p < .02$. As is obvious from Table 1, there was essentially no difference in RTs when the target phoneme was on the real vs the nonsense word, but there was a substantial difference when the target was after the real vs the nonsense word. RTs averaged about 100 msec longer in the latter condition.

Discussion

The results from Experiment I call into question the hypothesis that phoneme identification must always wait until after lexical access (or even an attempt at it). If the hypothesis was true, then nonsense words would have led to longer RTs relative to the real-word controls both when the target phoneme occurred on the nonsense word as well as when it occurred after it. But RTs were not elevated when the target was on the nonsense word. Apparently, phoneme identification need not always wait upon word identification and the phonological representation that then becomes available to the listener. Instead, listeners can identify target phonemes on the basis of the phonetic code that is developed independent of lexical access.

TABLE 1

Reaction Times (msec) from Experiment I

	Target position	
Target type	On	After
Word	475	525
Nonword	481	626

The fact that RTs were longer after nonsense words than after real words is consistent with the view than subjects had a difficult time determining where the nonsense word ended. This delay in the assignment of word boundaries means that the listeners could not rapidly determine whether a phonological segment meeting the target specification in other respects was word-initial.

Finally, reaction times were longer when the target occurred after the nonsense/real word than when it occurred on it. Part of the effect (the amount observed when the sentences contained only real words) may be due to stress differences between the items carrying the targets in the On vs the After position. It has been shown that phoneme-monitoring RTs are shorter when the target-bearing word is stressed relative to its surrounding words (Cutler & Foss, 1977). In Experiment I approximately one-third of the sentences had equal stress on the items in the two positions. Of the remaining sentences, approximately two-thirds had stress occurring on the item in the On position, where RTs were shorter. (Degree of stress was determined subjectively by listening to the tapes after the experiment. The number of items receiving stress in the On position varied somewhat between tapes.) The magnitude of the RT difference for word targets in the On vs the After position (50 msec) is similar to the magnitudes observed in Cutler and Foss' study.

Overall, then, Experiment I supports the contention that listeners can gain access to a phonetic code without having accessed the lexical item containing it. However, since the comparison of interest in this experiment involved nonsense items, it is possible that the above results are in some way artifactual. It could be argued that recognizing a segment of a nonword cannot bear directly on questions involving the processing of normal speech since there cannot be, in our terms, a stored representation of the nonword's phonological code (although there can presumably be a constructed representation; see footnote 1). Experiments II and III permit us to meet objections based upon the occurrence of nonwords in the first study.

EXPERIMENTS II AND III

Experiments II and III differ from Experiment I in that they employed only real words. In place of the lexical status variable (i.e., real word vs nonsense word), these two studies manipulated word frequency. That is, the target phoneme occurred on either a high- or low-frequency word, or immediately after it. In other respects the logic of the experiments was similar to that of Experiment I.

Morton (1969) and others have proposed that the time taken to retrieve a word from the mental lexicon is inversely related to the word's frequency of occurrence in the language. According to Morton's logogen model, each word is associated with a theoretical entity (the logogen) that has a particular threshold. The logogen accepts information from the senses and from other logogens (the context). Only when enough information has been received such that the threshold of the logogen has been exceeded is the word available to the perceiver. The word-frequency variable is represented in the model by its effects on the chronic threshold value of the logogen; frequently occurring words have a lower threshold than infrequent words. This means that frequent words require less input from the senses or the context before they are activated and their information becomes available. More rapid access for high-frequency words is the result.

If high-frequency words are accessed more rapidly than low-frequency words, and if lexical access is a prerequisite for identifying a phoneme, then it follows that phoneme-monitoring RTs should be shorter when the target phoneme is carried by a high-frequency word. This prediction seems quite clear. It contrasts with our expectations. According to our present view, listeners need not always retrieve the lexical item before responding to its initial phoneme. Therefore, we expect a result analogous to that found in Experiment I: namely, no difference in phoneme-monitoring RTs when the target occurs on the low- vs the high-frequency word. In contrast we do expect to see a significant difference when the target occurs after the low- vs the high-frequency items, the time being longer after the low-frequency words. Since listeners will take longer to access a low-frequency word, they will take somewhat longer to establish where the end of that word is. This provides a word-boundary problem similar to (though of course less than) that caused by the nonwords in Experiment I. When listeners take longer to determine that a phoneme is word-initial, they will take longer to respond to it if it is the target. Morton and Long would also expect to observe a difference between low- and high-frequency words when the target occurs right after them, so the views do not diverge at this point. Indeed, earlier research using the phoneme-monitoring paradigm (e.g., Foss, 1969) has observed frequency differences in this condition.

Method

Design and materials. Forty basic experimental sentences were constructed, each having four versions derived from the two independent variables: a sentence contained either a high- or a low-frequency subject noun (with similar meanings); crossed with this variable, either that noun or the verb immediately following it began with the target phoneme (On vs After conditions). Four material sets were constructed such that each basic sentence occurred in one of its four versions across the material sets. Each material set contained all of the basic sentences, ten from each of the four conditions. The experiment was, then, a 2(noun frequency: *high* vs *low*) × 2(target position: *on* the noun vs *after* it) × 4(material sets) factorial, with the first two factors being within subjects and the last being between subjects.

The high-frequency nouns had a mean frequency of occurrence of 125.5 ($SD = 80.8$), while the low-frequency nouns had a mean frequency of 3.9 (SD = 3.6). These figures are taken from the Kucera and Francis (1967) frequency count. The high- and low-frequency noun pairs were similar in meaning (e.g., *boy, brat; coffee, cocoa*) and were, overall, equated for number of syllables: the mean number of syllables for the 40 high-frequency nouns was 2.05 (SD = .96); for the 40 low-frequency nouns the mean was also 2.05 (SD = .81).

The nouns of interest were, in each sentence, the subject noun of the main clause. Experiment II differed from Experiment III in that the latter experiment had a prepositional or adverbial phrase prior to the main clause of each sentence. In Experiment II the main clause began the sentence. An example sentence with /t/ as the target phoneme on the high/low frequency noun, and /b/ as the target after the noun is:

(Yesterday afternoon,) the (*t*eacher/*t*utor) *b*orrowed the article from the reference library.

Eight instances of the phonemes /b,k,d,p,t/ served as targets for both the nouns and the verbs. In each sentence, the target nouns were singular and the target verbs were past tense. All target verbs had a frequency greater than ten.

In addition to the 40 experimental sentences, each material set contained 10 initial practice sentences, 20 filler sentences with the target phoneme on the object noun, and 10 sentences in which the specified target phoneme did not occur. In other respects these fillers were like the experimental sentences.

Subjects. The subjects in each of the two experiments were 40 undergraduate psychology students at the University of Texas at Austin who participated in the experiment in partial fulfillment of a course requirement. Within each of the experiments, 10 subjects were assigned to each of the four experimental tapes (material sets). No subject served in more than one of the experiments.

Procedure. The procedure used in these two experiments was similar to that of Experiment I. In these studies there was, of course, no mention of nonsense items since none occurred. The comprehension test was also similar in form to that used in Experiment I. Only those subjects who made six or fewer mistakes on the 26-item recognition test were included in the study. This led to a subject rejection rate of approximately one-third.

Results

The mean RTs in the four conditions of Experiment II are shown in Table 2; the means from Experiment III are presented in Table 3. In both cases 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 by a procedure suggested by Winer (1962). This typically resulted in replacing about 2% of the data.

TABLE 2

Reaction Times (msec) from Experiment II

	Target position	
Noun frequency	On noun	After noun
High	402	391
Low	411	438

TABLE 3

Reaction Times (msec) from Experiment III

	Target position	
Noun frequency	On noun	After noun
High	397	482
Low	403	514

Analysis of variance (by subjects) on the data from Experiment II showed that, overall, target position (On vs After) was not significant ($F < 1$), while word frequency was significant, $F(1,36) = 17.31, p < .001$. The interaction of these two variables was also significant, $F(1,36) = 12.64, p = .001$. The most important comparisons are those between the high- and low-frequency nouns when the target is on the noun, and between them when the target is after the noun. Accordingly, planned comparisons between high and low frequencies on, and high and low frequencies after, were carried out separately. When the target was on the noun, the frequency variable had no effect: the analysis by subjects had $F_1 = 1.12$, the analysis by items had $F_2 < 1$. In contrast, when the target was after the noun, the frequency variable was highly significant: $F_1(1,39) = 24.89, p < .001$; $F_2(1,39) = 12.53, p < .005$. The value of $minF'(1,70) = 8.33, p < .01$. Thus, the data from Experiment II showed that the frequency of the noun had an effect on the time to monitor for the initial segment of the word after it; the effect did not appear on the noun itself.

Analysis of variance (by subjects) on the data from Experiment III showed that, overall, target position (On vs After the noun) was highly significant, $F(1,36) = 146.15, p < .001$, as was the effect due to frequency, $F(1,36) = 11.06, p < .01$. The interaction of these two variables was also significant, $F(1,36) = 4.56, p = .04$. Again, the most important comparisons are those between the high- and low-frequency conditions when the target occurred on the noun and when it occurred after the noun. When the target occurred on the noun, the frequency of that noun had no effect on RTs; both the analysis by subjects and by items had $F < 1$. When the target occurred after the noun, however, the frequency of the noun had a significant effect: $F_1(1,39) = 11.25, p < .005$; $F_2(1,39) = 6.85, p < .02$. The value of $minF'(1,74) = 4.26, p < .05$. With the exception of the significant overall effect for target position (which is interesting but does not concern us here), the results from Experiment III closely replicated those from Experiment II.

Discussion

The data from Experiments II and III showed that the frequency of a subject noun had no effect on the time to respond to its initial phoneme.

That noun's frequency did, however, have a substantial (about 40 msec) and significant effect on the time to respond to the initial phoneme of the verb immediately following it. These results are conceptually identical to those from Experiment I. The status of the word carrying the target (nonsense item, low- or high-frequency word) has no effect on the time that subjects take to respond to the target. In contrast, the frequency of the item just prior to the target-carrying word has a significant impact upon the time to respond to the target.

The results from Experiments I–III are in sharp contrast to the views expressed by Foss and Swinney and by Morton and Long. These investigators assumed that phonological information is identified subsequent to lexical retrieval. They also assumed (correctly, we think) that lexical access is slower for low-frequency words than for high-frequency words. Consequently, listeners in Experiments II and III should have been able to respond more rapidly to the initial phoneme of the more rapidly retrieved (i.e., high-frequency) word. But this was not the pattern of data observed here.

The reasonable (almost necessary) interpretation of Experiments I–III is that listeners do not need to wait until lexical information has been retrieved in order to respond to word-initial target phonemes. Instead, they can respond (at least on occasion) as soon as a phonetic representation of the input has been developed by the speech processing mechanisms. This conclusion permits us to make a choice between the two broad classes of speech perception models described earlier. The data clearly favor models in which a representation of the input in terms of a phonetic code is developed prior to lexical access. Our subjects were responding to this code. The data argue against models in which no such representation is computed. Thus, they do not support the position taken by Klatt (1980) and Warren (1976), cited earlier.

In contrast, the reasonable (almost necessary) interpretation of the Morton and Long experiment is that listeners responded to the target phonemes subsequent to lexical access. We will shortly present the outlines of a theory that provides a resolution to the apparent conflict between the two interpretations. Before that, however, it is useful to compare the two sets of experiments to see where they differ. The most obvious difference between them is in the nature of the independent variables that were manipulated. In Experiments I–III the independent variable was intrinsic to the target-bearing word—its lexical status (Experiment I) or its frequency (Experiments II and III). In the studies conducted by Morton and Long the target-bearing words were equated for frequency but differed in their predictability. Unlike lexical status and frequency, predictability is determined by factors extrinsic to the target-bearing word itself; it is a function of the sentential context. Thus, in one sense, the relevant manipulation occurred prior to the target-bearing word in Morton and Long's study. In Experiments I–III the sentential context did not differentially favor one or the other of the target-bearing words. We will return momentarily to a discussion of this difference between the studies.

There was also a second notable difference between Experiments I–III and the work of Morton and Long. In their experiments the subjects were asked to recall verbatim each sentence immediately after it had been presented. In Experiments I–III the subjects were given the recognition comprehension tests described in the above Method sections. It is plausible to suppose that the processes involved in phoneme monitoring might be affected by the type of comprehension task given to the listeners (for a related point, see Triesman & Squire, 1974). Accordingly, it seemed advisable to determine whether the results obtained in the Morton and Long experiment would hold up when our comprehension task was used. Hence, we replicated the Morton and Long experiment, using a subset of their sentences while manipulating the type of comprehension task that was presented to the subjects. In Experiment IV some subjects were given the rote recall test while others were given the recognition test.

EXPERIMENT IV

In this experiment the contextual variable used by Morton and Long was manipulated, as was the type of comprehension test. Subjects were presented with two types of sentences: for any given sentence the target-bearing word was either relatively predictable or unpredictable. Half of the subjects were tested with the rote recall test used by Morton and Long; half were tested using the recognition test employed in the earlier studies in this series.

Method

Design and materials. Twenty basic experimental sentences taken from Morton and Long (1976) were used. Each sentence contained either a probable or an improbable noun as determined by the initial part of the sentence. For example, *He sat reading a book/bill until it was time to go home for his tea.* Probability of occurrence within the sentence was determined on the basis of norms collected by Morton. Within an experimental sentence the probable and improbable nouns were matched for initial phoneme and for frequency according to the Thorndike and Lorge (1944) estimates. For further details on the selection of probable and improbable nouns, see Morton and Long (1976). The target phoneme for an experimental sentence was the initial phoneme of the probable/improbable noun. Five stop consonants (/b,d,p,t,k/) were used as targets. The 20 experimental sentences are listed in the Appendix to Morton and Long's paper. Minor changes were made in a few sentences in order to eliminate some Britishisms.

Two material sets were constructed. Each set contained all 20 basic sentences, 10 of which occurred with the probable noun and 10 with the improbable noun. The materials were counterbalanced such that across the two sets each basic sentence occurred in both its probable and its improbable versions. Twenty-six filler sentences were also used. Ten did not contain the target that was specified for them; the remaining 16 had the targets occurring in various positions throughout them. The 46 sentences were randomized, with each basic sentence occurring in the same position in the two material sets.

Half of the subjects were tested using Morton and Long's rote recall test (Rote group); half were tested using the recognition test (Recognition group). The experiment was, then, a 2(contextual probability of occurrence of the noun: probable/improbable) × 2(comprehension test: rote/recognition) × 2(material sets) factorial, with the first variable within subjects and the last two variables between subjects.

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. Thirty subjects were tested with each comprehension task; half of the subjects within each of these groups received each material set.

Procedure. In most respects, the procedure used in Experiment IV was similar to that used in the prior experiments. Subjects were tested in groups of one to six people. The comprehension task was alternated between groups as the counterbalancing across material sets would permit. Subjects in the Rote group were given a numbered response sheet and told to write down each sentence immediately after they heard it. They were given 20 sec to record their responses. Subjects in the Recognition groups were forewarned in the instructions that a comprehension test would be administered after hearing all of the sentences. The intersentence interval for this group was only about 5 sec. The recognition test was constructed similarly to that used in the earlier experiments. It consisted of 24 sentences, half of which the subjects had actually heard. Again, the filler sentences provided the positive cases; some foils were superficially similar to other filler sentences. Only those subjects who scored above chance on the recognition task were included in the experiment.

Results and Discussion

The RT data were truncated as described in the Results section for Experiments II and III. The mean value of the RTs after probable and improbable nouns are shown in Table 4. The scores are highly similar across the two comprehension tasks. We will report the statistics for each task separately. For the Recognition group the difference between probable and improbable nouns was reliable, $F_1(1,29) = 18.99$, $p < .001$; $F_2(1,19) = 7.60$, $p < .03$; $minF'(1,34) = 5.43$, $p < .05$. Likewise, the difference in RTs to targets on probable and improbable nouns was significant for subjects in the Rote group, $F_1(1,29) = 31.19, p < .001$; $F_2(1,19) = 10.16, p < .005$; $minF'(1,31) = 7.66, p < .01$.

The main results reported by Morton and Long were observed for subjects given both types of comprehension task. Subjects responded more rapidly to targets carried by contextually predictable words than to

TABLE 4

Reaction Times (msec) from Experiment IV

	Noun type	
Comprehension test	Probable	Improbable
Recognition	405	449
Recall	409	456

targets carried by less predictable words. Thus, the major difference between the Morton and Long studies and Experiments I–III seems to reside in the nature of the variables manipulated and not in the type of comprehension tasks used. Apparently, experiments that manipulate transitional probability yield results consistent with the hypothesis that subjects respond to a target after retrieving the target-bearing word. In contrast, experiments that manipulate inherent characteristics of the target-bearing word (e.g., lexical status or frequency) yield results indicating that subjects respond prior to retrieving that word. Our account of this seeming discrepancy is perhaps best presented in the context of a view of speech perception from which it draws heavily.

GENERAL DISCUSSION

Earlier we noted that speech perception models differ with respect to the question of whether or not phonetic segments are computed during comprehension. Perceptual models also differ along another, orthogonal dimension. Some models emphasize the analyses that are carried out upon the acoustic signal (analytical or bottom-up models), while others emphasize the contribution that the listener's constructive mechanisms make to the perceptual process (synthetic or top-down models). Experiments implicating feature detectors in the speech perception process (e.g., Eimas & Corbitt, 1973) support the analytic perspective; experiments demonstrating the perception of speech segments in the absence of acoustic information (e.g., Warren & Obusek, 1971) support the synthetic view. Recent theorizing has led to the suggestion that both analytic and synthetic mechanisms are involved in normal speech recognition (e.g., Marslen-Wilson & Welsh, 1978). Thus, analytical mechanisms may be able to transform the acoustic representation of the input into a partially specified set of phonetic features, a set of candidate syllable boundaries, and a restricted set of stress indicators. The synthetic mechanisms may be used to "fill in" the phonetic segments so that a more completely specified phonological or lexical representation results. Information about the earlier analytic decisions, as well as information about earlier syntactic and semantic decisions may be used to guide these synthetic processes. The most successful artificially intelligent speech recognition programs (e.g., Reddy, 1976) operate with such mixed analytic/synthetic procedures. Indeed, it would be difficult to defend a pure model of either type in the face of the facts of speech perception.

To amplify a little, the mixed model of speech perception suggests that all of the codes mentioned earlier—acoustic, phonetic, lexical, and phonological—are psychologically realized during the course of comprehension. Further, it suggests that under certain circumstances some codes are more fully specified than others. To see, this, let us trace an input stimulus. The input is represented in an acoustic code under any circumstances. This is true even for utterances in an unknown language. Analytic processes operate upon the acoustic code producing a set of phonetic segments (or perhaps more accurately, a set of phonetic feature bundles). These processes are not limited to operating upon small stretches of the input. The analytic mechanisms must be sensitive to the dynamic aspects of speech, perhaps taking the syllable or half-syllable as their domain (e.g., Massaro, 1972; Studdert-Kennedy, 1976). It is often the case, however, that these analytic mechanisms are only partially successful (e.g., when speech is heard in a noisy environment). When that happens, the computed phonetic representation is only partially specified. Note that at this point the phonetic representations have not yet been identified as particular words. The process of lexical access utilizes the output of the analytic mechanisms; that is, it makes use of the phonetic code (as well as other information such as semantic context, expected part of speech, etc.).

When an item is retrieved from the mental lexicon, then all of the information stored with that item becomes potentially available to the listener. In particular, the complete phonological form of the word is made available. If there were any segments or features missing from the phonetic code they are now no longer missing. Thus, the phonetic code is often incomplete (some features required for identification of the phonetic segment may be missing), while the phonological code is completely specified.[2]

THE DUAL CODE HYPOTHESIS

We propose that both the phonetic and the phonological codes are developed during speech perception. In addition, we suggest that listeners can, under certain conditions, gain access to either of these codes and also make a response contingent upon a segment represented within either of them. Thus, in the phoneme monitoring task it makes sense to ask *which* phoneme is identified, i.e., does the subject respond on the basis of the phonetic or the phonological phoneme? Our answer is: either. When the phonetic code is specified completely enough, a response may be (though it need not be) made on the basis of it. When the phonetic code does not contain enough information to permit identification of the segment, then

[2] We are aware that this usage does not correspond to that preferred by such phonological theorists as Chomsky and Halle (1968). According to their framework, phonological segments are incompletely specified while phonetic segments are fully specified. Phonological rules operate upon the underlying phonological feature matrixes resulting in the surface phonetic matrixes. As is well known, their model does not purport to be a theory of perception. The present distinction is speaking to perceptual problems. Also, the disagreement is more apparent than real, a point we will not amplify upon here.

the response must be made on the basis of the phonological code if it is to be made at all. A related suggestion, using somewhat different terminology, has been independently made by Newman and Dell (1978) and by Cutler and Norris (in press).

This "Dual Code" hypothesis has at least two versions. In one, the subject has two distinct internal target representations, one phonetic and one phonological. These are loaded into phonetic and the phonological monitoring devices, respectively. Further, a threshold is associated with each monitoring device such that it will not report that a target has been found unless the input to the device exceeds the specified threshold. In the alternative view, there is a single abstract target entity, the phoneme, akin to a "phonological logogen" (cf. Morton, 1969). This is loaded into a device that accepts inputs from multiple sources (e.g., from the phonetic code, from the lexicon) and responds when its threshold has been exceeded. According to this version of the hypothesis, the listener does not respond to a phonetic code by matching it with an internally specified phonetic target; rather, the phonetic code is sometimes sufficient by itself to trigger the abstract phoneme detector.

A further refinement of the present theory is required if we are to understand how monitoring works. We view the monitoring task as one that requires active attention on the part of the listener; it certainly is not an autonomous process in the way that the phonetic analysis itself is. Of course, processing the sentence also requires the listener's resources. As a result of these multiple demands upon the listener, queuing problems can arise. Scheduling algorithms must be devised in order to handle the ongoing tasks in accordance with the payoffs for completing them.

The experiments under discussion can be accounted for by the Dual Code hypothesis. In Experiments I–III the subjects were typically responding on the basis of the phonetic code. They did not have to gain access to their lexicons in order to determine the phonological code before they were able to initiate their responses to the target phoneme. In contrast, the subjects in Morton and Long's experiment and in Experiment IV were typically responding to the phonological code. Lexical access occurred prior to response initiation. But why should this difference occur? Recall that according to the Dual Code hypothesis subjects can respond to either the phonetic or the phonological code. As soon as one of the monitors exceeds its threshold a response can be initiated—assuming that the scheduling algorithm has the monitoring task as its first priority. It is a probabalistic matter as to which monitor will first exceed its threshold in any given sentence. The probabilities are not fixed, however. Various factors can alter the likelihood that the response will be made to the phonetic or to the phonological codes.

Earlier we noted that there are two major differences between Experiments I–III and the study conducted by Morton and Long. One difference is the type of comprehension test used. This appears to have little effect on the code to which listeners respond (but vide infra). The second and more important difference is the nature of the independent variables in the two sets of studies. In Experiments I–III the experimental manipulation occurred on the target-bearing word, while in the Morton and Long studies (and their replication here) the manipulation occurred prior to the target-bearing word. How could this difference affect the probabilities of responding to the phonetic or the phonological codes?

Consider again the course of events during speech perception. The acoustic representation is operated upon by a set of stimulus analyzing mechanisms. We conjectured that the output of these mechanisms is a (partially specified) set of phonetic feature matrices. If one of the segments is preceded by a word boundary, and if this segment shares enough features with the phonetic target, then the phonetic monitoring device will detect it. A response can be initiated at that point if monitoring has high priority. The phonetic representation also provides the major source of information for lexical access. Of course, this is its primary role in the language system; the monitoring task is an unusual and attention-demanding appendage.

It is plausible to assume that the phonetic representation will often be imperfect. That is, there may be missing entries in the phonetic feature matrix. This may occur in noisy environments, to take the most obvious example. An incomplete phonetic representation may have various consequences. Sometimes a failure of lexical access will occur. At other times contextual information may be sufficient to lead to lexical access even when the phonetic code is imperfect. In particular, this is likely to happen when words occurring earlier in the sentence are semantically related to the appropriate lexical entry. Blank and Foss (1978) have shown, for example, that access to a word in a sentence is speeded when that word is preceded by a semantically related word. In some cases, then, the phonetic specification of a segment may not be sufficient to cause the phonetic monitor to become activated. At the same time, the incomplete phonetic specification of, say, the initial syllable, plus the semantic context, may be sufficient to activate a particular entry in lexical memory. When this happens the phonological specification of the word becomes available and the subject's phonological monitor will register that a target has occurred. What is interesting about this case is the fact that the phonological code may be sufficient to lead to a response while the phonetic code is still being developed and is not yet sufficient to activate responding.

We can account for Morton and Long's results by assuming that information from the phonological code of expected (i.e., high transitional probability) target-bearing words was available and used for identification of the target segment before sufficient information derived from the phonetic code could lead to identification of the target. In this case, then, subjects were responding to a code that arose subsequent to lexical access; in consequence, one observes the effects of lexical access speed on phoneme monitoring RTs. In the case of Experiments I–III, however, the listeners were not able to use context to differentially activate the target-bearing lexical entries; relatively complete phonetic specifications were required for lexical access to occur. And by the time the phonetic code was sufficiently developed so that lexical access could occur, it was also sufficient for identification of the target segment by the phonetic monitoring device. In this case, then, one observes no effects of lexical access speed in the phoneme monitoring task.

Summarizing to this point, we have proposed that both the phonetic and the phonological codes are computed by the speech processor and that responding to either of these codes requires attention. In addition, we have suggested that there may be competition for resources between the monitoring task and the task of comprehension itself. And we have also argued that the code to which listeners respond in the phoneme monitoring task can be manipulated by varying such parameters as target-word predictability. Obviously these ideas need to be made more precise and testable. There is in the existing literature, however, other evidence that supports the Dual Code hypothesis.

Related Support for the Hypothesis

In this section we will use the Dual Code hypothesis to guide us as we examine the relationship between the monitoring task and other aspects of comprehension. Our aims are to explore the conditions under which one or the other of the two codes is likely to be responded to, and to clarify the model itself. We will consider briefly three sources of data: effects of memory load on listeners due to the number of targets that they are monitoring for, effects due to phonetic similarity between the target item and other items in the sentence, and effects of the comprehension task used in the experiments.

Memory load. While the Dual Code hypothesis can account for the existing data obtained using phoneme monitoring during sentence comprehension, there is in the literature a study employing word lists that does not appear to fit readily into this framework. Rubin, Turvey, and Van Gelder (1976) reported the results of two experiments in which subjects were asked to monitor for word-initial target phonemes. In their studies subjects were presented with lists of monosyllabic words and nonwords. They found that RT was significantly shorter when the target phoneme was carried by a word than when it was on a nonword. It would appear that the Dual Code model predicts that subjects should respond more often to the phonetic than to the phonological code in this study (it bears a resemblance to the present Experiment I). Therefore, RTs should not have differed between the word and nonword targets. The finding that an item's status as a word or a nonword did have an affect on RT to that item's initial phoneme is embarrassing for the Dual Code hypothesis.

The response to this embarrassment can be more than blush, however. There are a number of reasons why the particular experiment cited might have come up with the results it did. Among the prominent differences between the Rubin et al. study and Experiment I is the fact that subjects in the former were asked to monitor for two targets (/b/ and /s/). Thus, they were making a choice response, pushing one button if a stimulus item began with /b/ and another button if it began with /s/. This small difference in the task might make a very big difference in the probability that the phonetic or the phonological code will be responded to first.

It has been shown that monitoring for two or more targets puts an additional load on the subjects in a phoneme monitoring task such that RTs increase (Foss & Dowell, 1971). Presumably, this increased processing load results from having to test phonemes for multiple sets of attributes. It seems plausible that the processing demands on the subjects due to the number of targets might affect the ease with which they can gain access to one or the other of the two codes of interest. The plausibility of this assumption follows from another one, namely, that the phonetic code is more transient than is the phonological code. The phonetic code is used to access entities in the mental lexicon. Since this is typically a very rapid process, there is no reason for the phonetic code to stay active for very long. Once a lexical entry has been accessed, its phonological code becomes available. Since phonological codes are derived from a lexical entry, and since lexical units are used by the syntactic and semantic processors, phonological codes are likely to be available long after the phonetic code has "faded." This suggests, then, that subjects have a rather narrow time window during which they can respond to the target on the basis of the phonetic code. The time window is much wider for the phonological code.

Returning to the results observed by Rubin et al. we propose that the presence of two targets lowered the probability that the listeners were able to examine the transient phonetic code before it faded. This is tantamount to predicting an increase in the probability that the subjects responded to the phonological code. Consequently, the modified Dual Code hypothesis predicts that one will observe an effect due to the status of the target-bearing item (word vs nonword) in the Rubin et al. study.

The present analysis is corroborated by the results from an experiment reported by Rubin (Note 1). He carried out a phoneme monitoring study in which subjects were presented with lists of monosyllabic words and nonwords. Unlike the Rubin et al. study, however, subjects in Rubin's experiment were instructed to monitor for a single target. Under these circumstances the Rubin et al. effect went away. There was no difference in RTs to word-initial targets on words and nonwords; the RTs were 652 and 660 msec, respectively. Since lexical status did not affect RTs, we must assume that the subjects in Rubin's experiment were able to respond to the phonetic code, just as were the subjects in Experiment I. Thus, when task demands preclude monitoring at the phonetic level we will observe the effects of variables that affect the time required to access the target-bearing item.

Phonetic similarity. Newman and Dell (1978) found that phoneme monitoring latencies are affected by the phonetic similarity between the target phoneme and the initial phoneme of the word immediately preceding it (the "critical" phoneme). Specifically, as the number of shared distinctive features between the critical and target phonemes increases, so does the RT to the target. Newman and Dell took these findings as supporting "a role for a bottom-up procedure in phoneme identification, in which at least part of the identification is carried out directly via the acoutic properties of the stimuli" (p. 371).

Newman and Dell's finding permits additional insights into the conditions under which phonemes are responded to phonetically vs phonologically. When a listener in a phoneme monitoring study encounters a word-initial phoneme that is similar to the specified target, the monitoring device will have a tendency to respond, and therefore resources will be devoted to this part of the input. A more complete analysis of the critical segment may be instituted in order to avoid false alarms. As long as resources are devoted to analyzing the critical phoneme, less attention can be paid to the next part of the sentence—the point that actually contains the target. If the phonetic code of the actual target is not examined quickly, it will fade. In that case, of course, the listener will not be able to respond to the target on the basis of this code. Identification of the target phoneme will have to occur on the basis of the phonological code. In contrast, when the actual target is preceded by a dissimilar critical phoneme, there is no special increase in demand for processing resources. Inspection of the phonetic code of the target phoneme will be possible and subjects will be able to respond to it.

The Dual Code analysis of the effects due to phonological similarity thus generates the following prediction. Target phonemes preceded by similar critical phonemes should be responded to postlexically (i.e., phonologically). Under these conditions, phoneme-monitoring RTs should reflect variables such as semantic relatedness that presumably affect speed of lexical access. On the other hand, target phonemes that are preceded by dissimilar critical phonemes should be responded to prelexically (i.e., phonetically). The effect of such variables as semantic relatedness should be minimal under these conditions.

In recent work (Note 2), Dell and Newman presented subjects with some sentences in which prior context was predictive of the target-bearing word and some in which it was not. They also manipulated, in a between-subjects design, the phonological similarity of the critical and target phonemes. Dell and Newman correctly noted that the Dual Code hypothesis predicts an interaction between the context and the phonological similarity variables in their study. And, indeed, they observed the predicted interaction. Subjects who were presented with critical phonemes similar to the target phoneme showed a large (c. 100 msec) effect of semantic relatedness. This replicates the Morton and Long study and Experiment IV. Subjects who had critical phonemes dissimilar to the target phoneme did not show the effect (c. 15 msec). The interaction was highly reliable. Thus, Dell and Newman have shown that one can manipulate the code, pre- vs postlexical, to which subjects will respond. The phonological similarity variable controls the probability of responding to the phonetic code by affecting the allocation of processing resources. This result both supports the Dual Code model and increases our understanding of the variables that control the code to which subjects respond.

Comprehension tasks. We have noted that subjects in monitoring experiments have many demands placed upon them and that they probably must deal with these demands by setting priorities (the problem of the scheduling algorithm). It seems plausible that the type of comprehension task given to subjects might affect the ease with which they can gain access to one or the other of the two codes of interest. Suppose that a relatively difficult comprehension task is presented and that subjects are paid off for completing it (i.e., comprehension is first in the queue for resources). In this case, the chances that subjects can respond to the phonetic code should be decreased. By the time the demands of comprehension have been attended to, the phonetic code will have faded. Listeners must then respond to the phonological code if they are to respond at all. Thus, if we can manipulate the degree to which listeners attend to the task of understanding the sentence, we will also be manipulating the degree to which they are able to respond to the phonetic code.

In Experiment IV we manipulated the type of comprehension task that was presented to the subjects, so on the basis of the present analysis we might expect that there would have been differences in the probability of responding to the phonetic code as a function of the type of task that was used. One problem with this analysis, however, is that we do not have a very good idea about which of the two comprehension tasks (rote recall vs recognition) puts the greater demand upon the sentence comprehension mechanisms at the point where the target phoneme occurred. Our theory of comprehension tasks is not well enough developed to make this choice, and armchair analyses are often inadequate (see Britton, Westbrook, & Holdredge, 1978, for an interesting discussion of such a problem). Clearly what is needed are experiments that manipulate the degree to which subjects put the comprehension task at the front of the attention queue. This can perhaps be most effectively done by manipulating payoffs for good performance on the monitoring and comprehension tasks.

Dual Coding and Other "On-Line" Tasks

To this point we have described the outlines of the Dual Code model and have seen how it can help us to understand the results of several phoneme monitoring experiments. The model also has relevance for interpretations of data gathered with other tasks. In fact, it can help us integrate in a coherent way results obtained from other paradigms used to investigate the perception of phonemes in fluent speech. At the same time, data gathered using these tasks may help us to define more explicitly the model itself. In this section we will describe briefly the results of some experiments using speech shadowing, mispronunciation detection, and phonemic restoration, and we will see how these data are related to the model.

Shadowing. In the shadowing task listeners are aurally presented with sentences and are asked to repeat back what they hear as rapidly as they can. Important data exploring the mechanisms of shadowing have been gathered by Carey (1971), Marslen-Wilson (1973), and Miller and Isard (1963), among others. Here we focus upon results from a shadowing study in which some of the words presented to the shadower were deliberately mispronounced (Marslen-Wilson & Welsh, 1978). When presented with such an input, listeners either can give back an exact repetition (i.e., repeat back the mispronunciation) or they can restore the input to its "intended" form such that they produce the speech without the mispronunciation. Marslen-Wilson and Welsh found that exact repetitions typically were associated with other disfluencies in shadowing, notably pauses. This suggests that the listeners were aware of the mispronunciations and that their syntactic and semantic analyses were disrupted by the deviant input. On the other hand, the restorations were typically fluent ones, no other disfluencies occurred in the subjects' speech. This suggests that such restorations reflect true nonperception of the mispronunciations. That is, the restorations were apparently made without the subjects' knowledge that they had made a change in the input. If the subjects had actually perceived the error and had then made a conscious correction of it, we would expect to observe disfluencies in the restorations.

In terms of the Dual Code model, we propose that fluent restorations reflect a state of affairs in which the listeners actually "heard" the phonological code associated with the intended word, i.e., they heard the code that is stored in the mental lexicon. This point of view is similar to that forwarded by Chomsky and Halle (1968), who argued that listeners hear the underlying phonological representation of the input rather than its phonetic representation. In addition, we propose that when subjects exactly repeated the error, they "heard" what was represented in the stimulus, i.e., they heard the phonetic code. The associated disfluencies arose either because the listeners also retrieved the appropriate lexical item and noted the discrepancy between critical features in the two codes, or because higher level analyses were disrupted by the deviant input. This latter position, the claim that listeners have the ability to hear phonetic codes directly, is somewhat different from that suggested by Chomsky and Halle.

The Dual Code Model does more than provide labels for the codes to which subjects are responding when they make these two types of responses, however. In particular, the model enables us to predict when listeners will make an exact repetition and when they will make a fluent restoration. Marslen-Wilson and Welsh manipulated some parameters that affected the relative likelihood of these two types of shadowing responses. They found that the probability of observing a fluent restoration is greater when the prior context of the sentence containing the mispronunciation is semantically related to the word mispronouned. And for sentences with semantically related contexts, they also observed more fluent restorations when the mispronunciation occurred within a word than when it occurred at the beginning of a word. These observations are predicted by the analysis given earlier when we discussed the results of the Morton and Long study and Experiment IV.

Recall that according to the Dual Code hypothesis subjects are more likely to gain rapid access to a word's phonological code when prior context is semantically related to that word. A related context permits the word to be accessed with relatively small amounts of phonetic information. Therefore, listeners will often have available to them the (intended) phonological code before sufficient phonetic information is analyzed to permit identification of the mispronunciation. In such cases listeners will "hear" the phonological code and a fluent restoration will result, the subjects repeating back what was heard. The Dual Code model also predicts that the phonological code is more likely to be perceived when mispronunciations occur late in a word. In such cases the chances are high that the intended word will be accessed on the basis of prior context and the word's early phonetic code. Deviations from correct pronunciation will often go undetected since the listener "hears" the phonological code as soon as the word is accessed. In contrast, when the phonetic code of the initial part of the word is mispronounced, lexical access of the intended word will be impaired. Consequently, listeners will be more likely to hear the phonetic code and to detect the fact that it deviates from the intended word when that word is finally accessed. (See Marslen-Wilson, 1973, 1975 for related work that is consistent with this interpretation.) In sum, we interpret this work as showing that syntactic, semantic, and intraword constraints influence whether listeners perceive the phonetic or the phonological codes. The Dual Code model, along with the assumption that listeners' experiences are partially determined by which code is available, permit us to predict these shadowing results.

Mispronunciation detection. In this task listeners are asked to detect a mispronunciation that is usually presented in a sentence context. The mispronunciation detection task is like phoneme monitoring (and unlike shadowing) in that it focuses subjects' attention on the sound structure of the utterance. According to Cole (1973), subjects can detect a mispronunciation in one of two ways: they can discover that the phonetic code of the input does not match any item in the mental lexicon, thus leading them to conclude that it must be a mispronunciation; or they may access a word via the phonetic code and detect a mismatch between critical features of this code and the stored phonological code. There is good evidence that the latter is the correct description of what occurs in the typical experiment (Cole & Jakimik, 1979).

Cole and his associates (1973; Cole & Jakimik, 1979) have demonstrated that the probability of detecting a mispronunciation is a function of the number of distinctive features by which the mispronounced segment of the input item differs from the intended word. Of more interest to us here, however, is the fact that the probability of detecting a small (one feature) discrepancy is greater when it occurs at the beginning of the word's first syllable than when it occurs in the second or third syllable of the word (Cole, 1973; Marslen-Wilson & Welsh, 1978). This result is quite parallel to the result observed in the shadowing studies, where the probability of an exact repetition of a mispronounced word (and of attendant disfluencies such as pauses) was greater when the mispronunciation occurred early in the word. Our explanation for this observation is parallel to that given for the shadowing data: when the mispronunciation occurs late in a word, subjects are likely to have already accessed the appropriate item in the mental lexicon and to "hear" the phonological code. Consequently, they will be unlikely to detect a small discrepancy between that code and the phonetic code.

The above explanation of the asymmetry of detecting early vs late mispronunciations enables us to interpret sensibly a finding reported by Cole, Jakimik, and Cooper (1978). These investigators manipulated both the place within the word where the mispronunciation occurred (initial vs final consonant), and the type of feature that was changed (voicing vs nasality). They found that mispronunciations were detected more often when they occurred in word-initial position than when they were in word-final position. Cole et al. attempted to explain this effect in terms of acoustic factors: "It seems likely that voicing differences are more perceptible in word-initial than word-final stops because the acoustic cues for voicing are more stable and well defined in word-initial position" (p. 55). However, the authors found that the positional effects observed for nasals could not be explained in a similar way: "On the other hand, we are unaware of any data which show that acoustic cues for /m/–/n/ distinctions are more salient in word-initial than word-final position" (p. 55). Consequently, they suggested that the asymmetry could be accounted for by the hypothesis that listeners pay more attention to the beginnings of words than to their later sounds.

According to the Dual Code model, increased sensitivity to mispronunciations in word-initial position is due to the fact that initial sounds are perceived on the basis of their phonetic codes. In contrast, mispronunciations are less likely to be detected in word-final positions because the perception of the later segments is most likely mediated by stored phonological representations. Thus, the model accounts naturally for the results observed by Cole et al.

Phonemic restorations. The Dual Code model is also relevant to the body of research concerned with the phenomenon known as phonemic restoration. And, as with previous data, the phonemic restoration effect can help us to understand and further test the model. Warren (1970; Warren & Obusek, 1971) first demonstrated the restoration effect. He found that when part of the speech waveform is deleted and replaced with noise, listeners report hearing the missing speech sound(s) as clearly as the segments that are physically represented in the input signal. Obviously, a phonetic code which is computed via an analysis of the acoustic signal cannot mediate the perception of phonemes that are not actually represented in some portion of the speech signal. Thus, the perception of "restored" phonemes is necessarily based on the phonological code, while the perception of "actual" phonemes may be based upon either the phonetic or the phonological codes, as noted in earlier sections.

According to the Dual Code model, the quality of the illusory restoration should depend in part on the degree to which context permits lexical access on the basis of partial phonetic information. To our knowledge parametric tests of the appropriate sort have not been carried out. However, the results of a phonemic restoration experiment conducted by Sherman (Note 3) corroborate the model's claim that semantic context affects the availability of phonological codes. In this study, listeners heard sentences such as *It was found that the *eel was on the* ____, in which the last word was varied (the * indicates the noise). The last word was either *axle, shoe, orange,* or *table.* These words provide contexts which could affect the particular phoneme that would be restored. The restorations related to the above words are *wheel, heel, peel,* and *meal,* respectively. Sherman found that listeners heard the phoneme appropriate to the semantic context, even though this context occurred after the deleted segment. (The extent to which the context can be delayed, and the extent to which these restorations are as complete as those occurring when the context comes before the missing segment, are both undetermined to date). It is consistent with the Dual Code model that subsequent context could affect which word is accessed and therefore which phonological code becomes available to listeners. However, the model predicts that restorations would be less frequent in this case than with prior context since the phonetic code may fade before subsequent context is able to aid in the access of the related lexical item.

The existence of the phonemic restoration effect provides a fertile field for testing further the Dual Code model of perception. Combining the phonemic restoration phenomenon with the phoneme monitoring task permits one to generate some interesting predictions. (Sherman, Note 3, has demonstrated that subjects can monitor for restored phonemes.) To mention just one, the time required to respond to a restored phoneme should reflect the time needed to access the word carrying that phoneme (i.e., response time should reflect the time needed to gain access to the phonological code). Thus, when the target phoneme must be restored one should observe a different pattern of results than that obtained in Experiments I–III. Recall that in those studies word frequency or lexical status had no effect on RTs when the target was on the high- vs the low-frequency word (or the nonword). According to the Dual Code hypothesis, these data reflect responses to the phonetic code. However, if the target is excised so that it must be restored, then subjects must respond to the phonological code. In that case, then, a variable like frequency will have an effect on RT even though the target is on the high- vs low-frequency word. This prediction, along with a number of others, has been tested by Blank (Note 4). She found considerable evidence consistent with the Dual Code hypothesis.

SUMMARY

This paper has made five major points. First it was shown that subjects can respond to target phonemes prior to retrieving the words that carry the target segment. Second, we argued from these results that certain classes of speech decoding models—those that deny the psychological validity of the phonetic code—are untenable. We then proposed that both phonetic and phonological codes have psychological validity, and that subjects carrying out the phoneme monitoring task sometimes respond to one of these codes and sometimes to the other (we dubbed this the Dual Code hypothesis). Fourth, we discussed the Dual Code hypothesis, specifying some of the conditions that are likely to favor a response to the phonetic as opposed to the phonological code. Finally, the Dual Code hypothesis was used to integrate and clarify data obtained from a number of experiments using a variety of on-line measures of speech processing.

REFERENCES

Blank, M. A., & Foss, D. J. Semantic facilitation and lexical access during sentence processing. *Memory & Cognition*, 1978, 6, 644–652.

Britton, B. K., Westbrook, R. D., & Holdredge, T. S. Reading and cognitive capacity usage: Effects of text difficulty. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, 4, 582–591.

Carey, P. W. Verbal retention after shadowing and after listening. *Perception & Psychophysics*, 1971, 9, 79–83.

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Cole, R. A. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 1973, 13, 153–156.

Cole, R. A., & Jakimik, J. Understanding speech: How words are heard. In G. Underwood (Ed.), *Information processing strategies*. New York: Academic Press, 1979

Cole, R. A., Jakimik, J., & Cooper, W. E. Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 1978, 64, 44–56.

Cutler, A., & Foss, D. J. On the role of sentence stress in sentence processing. *Language & Speech*, 1977, 20, 1–10.

Cutler, A., & Norris, D. Monitoring sentence comprehension. In W. E. Cooper and E. C. T. Walker (Eds.), *Sentence processing: Studies in honor of Merrill Garrett* (in press).

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, 4, 99–109.

Foss, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 457–462.

Foss, D. J. On the time-course of sentence comprehension. In *Problemes actuels en psycholinguistique/Current problems in psycholinguistics*. Paris: Editions du C.N.R.S., 1974.

Foss, D. J., & Dowell, B. E. High-speed memory retrieval with auditorily presented stimuli. *Perception & Psychophysics*, 1971, 9, 465–468.

Foss, D. J., & Lynch, R. H., Jr. Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception & Psychophysics*, 1969, 5, 145–148.

Foss, D. J., & Swinney, D. A. On the psychological reality of the phoneme: Perception, identification, and consciousness. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 246–257.

Klatt, D. H. Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. Cole (Ed.), *Perception and production of fluent speech*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Kucera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence: Brown University Press, 1967.

Marslen-Wilson, W. D. Linguistic structure and speech shadowing at very short latencies. *Nature (London)* 1973, 244, 522–523.

Marslen-Wilson, W. D. Sentence perception as an interactive parallel process. *Science*, 1975, **189**, 226–227.

Marslen-Wilson, W. D., & Welsh, A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 1978, **10**, 29–63.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124–145.

Massaro, D. W. Perceptual units in speech recognition. *Journal of Experimental Psychology*, 1974, **102**, 199–208.

Miller, G. A., & Isard, S. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 217–228.

Newman, J. E., & Dell, G. S. The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17**, 359–374.

Reddy, D. R. Speech recognition by machine: A review. *Proceedings of the IEEE*, 1976, **64**, 501–531.

Rubin, P., Turvey, M. T., & Van Gelder, P. Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 1976, **19**, 394–398.

Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. XII). The Hague: Mouton, 1974.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976. Pp. 243–293.

Studdert-Kennedy, M., Shankweiler, D., & Pisoni, D. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*, 1972, **3**, 455–466.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. N.Y.: Teachers College, Columbia University, 1944.

Treisman, A., & Squire, R. Listening to speech at two levels at once. *Quarterly Journal of Experimental Psychology*, 1974, **26**, 82–97.

Warren, R. M. Perceptual restoration of missing speech sounds. *Science*, 1970, **167**, 393–395.

Warren, R. M. Auditory illusions and perceptual processes. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*, New York: Academic Press, 1976.

Warren, R. M., & Obusek, D. J. Speech perception and phonemic restorations. *Perception & Psychophysics*, 1971, **9**, 358–362.

Wood, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 3–20.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

REFERENCE NOTES

1. Rubin, P. E. *Semantic influences on phonetic identification and lexical decision*. Unpublished doctoral dissertation, University of Connecticut, 1975.
2. Dell, G. S., & Newman, J. E. *The interactive nature of phoneme monitoring*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, 1978.
3. Sherman, G. L. *Studies of the temporal sequence of speech perception at different linquistic levels*. Unpublished doctoral dissertation, University of Wisconsin-Milwaukee, 1973.
4. Blank, M. A. *Dual-mode processing of phonemes in fluent speech*. Unpublished doctoral dissertation, University of Texas at Austin, 1979.

(Accepted June 18, 1979)