Perception & Psychophysics
1988, 43 (6), 541-550

The relative accessibility of phonemes and syllables

DENNIS NORRIS and ANNE CUTLER
MRC Applied Psychology Unit, Cambridge, England

Previous research comparing detection times for syllables and for phonemes has consistently found that syllables are responded to faster than phonemes. This finding poses theoretical problems for strictly hierarchical models of speech recognition, in which smaller units should be able to be identified faster than larger units. However, inspection of the characteristics of previous experiments' stimuli reveals that subjects have been able to respond to syllables on the basis of only a partial analysis of the stimulus. In the present experiment, five groups of subjects listened to identical stimulus material. Phoneme and syllable monitoring under standard conditions was compared with monitoring under conditions in which near matches of target and stimulus occurred on no-response trials. In the latter case, when subjects were forced to analyze each stimulus fully, phonemes were detected faster than syllables.

Speech recognition involves the matching of spoken word forms to lexical representations. In principle, this could be achieved by a simple exhaustive template-matching process. However, the number of potential lexical representations to be checked, and the difficulty of determining, in a continuous speech signal, where word forms begin and end, suggests that simple template matching would be a relatively cumbersome method of accessing lexical representations. Greater efficiency would be achieved with a preliminary classification of the speech signal, using a relatively small set of units of which any word form will be composed. Such a classification would greatly simplify the lexical access process, because exhaustive search of all representations would not be necessary. If the stored representations were arranged in an order determined by the units of classification, the preliminary analysis would allow the lexical forms to be accessed directly, in just the way that alphabetic arrangement of a dictionary allows an entry to be found without uncertainty.

For this reason, psycholinguists have expended considerable effort on investigating whether there are such "units of speech perception." The main candidates have been the syllable and the phoneme. The phoneme has the advantage of being the smallest unit into which speech can be sequentially analyzed. The syllable has the advantage of being the smallest spoken unit (with the possible exception of utterances composed only of hisses or hums). The experimental evidence to date, especially from reaction-time studies, seems to favor the syllable. Many studies have compared detection time for phoneme and syllable targets, and have consistently found syllables to be identified more rapidly than phonemes (Foss & Swinney, 1973; Mills, 1980b; Savin & Bever, 1970; Segui, Frauenfelder, & Mehler, 1981; Swinney & Prather, 1980).

Two studies, it is true, have suggested that phonemes can be identified faster than syllables under certain conditions. But in each of these studies, it is arguable that the subject's task in the phoneme-monitoring condition has in fact amounted to syllable monitoring. For example, McNeill and Lindig (1973) had subjects monitor for consonant targets in a condition in which the consonants were always followed by the vowel /a/. In other words, the targets were actually syllables whose exact form could be determined from the target specification. Healy and Cutting (1976) also reported a phoneme advantage under some conditions of their experiments. In this case, the targets were isolated vowels. Since vowels in isolation are effectively syllables, this experiment also amounts to a test of syllable monitoring with shorter versus longer targets, rather than of syllable monitoring versus phoneme monitoring.

The finding that syllables can be identified faster than phonemes therefore seems to be robust. This apparently contradicts the simplest kind of hierarchically structured perceptual system in which lower level units are perceived first and then combined into larger units. Moreover, whole-word targets can be detected even faster than can syllable targets (Foss & Swinney, 1973), which appears to argue against any kind of sublexical classification at all. Foss and Swinney attempted to resolve this paradox by drawing a distinction between perception and identifi-

Our interest in syllabic versus phonemic segmentation has been stimulated and nurtured by our collaboration on cross-linguistic research with Jacques Mehler and Juan Segui. We are very grateful to them for continuing discussion of these issues. Financial support for this research was provided by grants from British Telecom and from the European Training Programme in Brain and Behavioral Sciences of the European Science Foundation. Acknowledgment is made to the Director of Research of British Telecom for permission to publish the paper. We thank Donald J. Foss for providing us with a list of the materials used by Foss and Swinney (1973). We also thank Steve Bartram, Sally Butterfield, and John Williams for technical assistance, Ian Nimmo-Smith for statistical advice, and Phil Johnson-Laird for helpful comments on the manuscript. Address correspondence to D. G. Norris, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, U.K.

cation. They argued that although lower level units may actually be perceived before higher level units, perception may not automatically lead to awareness and identification. Furthermore, the order in which units are identified may not correspond to the order in which they are perceived. Indeed, the order of identification may be precisely the reverse of the order of perception. For instance, phonemes may be perceived before syllables, but syllables may become available for identification before phonemes. Since monitoring tasks presumably reflect the order of identification rather than the order of perception, it should not be surprising to find that words can be detected faster than syllables, which in turn can be detected faster than phonemes.

Foss and Swinney's (1973) distinction, however, seems somewhat counterintuitive. There is a large difference in average duration between words, syllables, and phonemes. All of the information necessary to identify the initial phoneme of a CVC syllable comes in the first half of the syllable. All of the information necessary to identify the first syllable of a two-syllable word comes in roughly the first half of the word. However, despite these large differences in length, the phoneme somehow takes longer to identify than the word. Although Foss and Swinney have constructed an argument whereby it is certainly logically possible for the phoneme to be a primary unit of perception, as long as the response-time differences stubbornly continue to favor the syllable and the word, the argument remains, from a theoretical standpoint, unsatisfying.

Moreover, there exists some direct evidence suggesting that the syllable can function as a basic perceptual unit. Mehler, Dommergues, Frauenfelder, and Segui (1981) showed that syllable-monitoring responses were facilitated when the target specification matched the syllabification of the target word. For example, although the words *balance* and *balcon* both begin with the same three sounds, the first syllable of *balance* is *ba-*, whereas the first syllable of *balcon* is *bal-*. Mehler et al. found that the target *ba* was identified faster than the target *bal* in *balance*, whereas in *balcon* the converse was true: *bal* was detected faster than *ba*. They argued that this result reflected listeners' segmentation of the speech input into syllables. When the target specification matched the segmentation, responses were faster.

It might be objected that Mehler et al.'s (1981) result is explicable in terms of a perceptual match between target specification and target. Mills (1980b) and Swinney and Prather (1980) have demonstrated that targets that more closely accord with the listener's expectancies about how the target will sound are identified more rapidly. In Mehler et al.'s experiment, then, subjects presented with the target specification *ba* might have simply converted it into an internal representation that was a better perceptual match to the beginning of *balance* than to the beginning of *balcon*. If this were the case, then Mehler et al.'s finding could of course be accounted for without reference to syllables at all. However, the perceptual-match hypothesis cannot explain why this effect should hold for French listeners but not for English listeners (Cutler, Mehler, Norris, & Segui, 1983, 1986). Cutler et al repeated Mehler et al.'s experiment with English listeners, presenting both an English version of the materials and the original French stimuli. In both cases, the English listeners were very little influenced by whether the target specification was CV (e.g., *ba*) or CVC (e.g., *bal*). The main factor influencing the English listeners was the structure of the target word itself; responses were faster to words like *balance* than to words like *balcony*.

As a possible explanation for this latter finding, Cutler et al. (1986) suggested that English listeners might have relied on a phonemic rather than a syllabic segmentation strategy. The difference between the different types of word, they argued, might be due to some sequences of phonemes being easier to perceive than others. Specifically, they proposed that consonants may be easier to perceive in the context of vowels, and vowels may be easier to perceive in the context of consonants. Both of these factors would act to make words like *balance*, which begin with a CVCV sequence, easier to perceive than words like *balcony*, which begin CVCC.

The cross-linguistic studies would seem to suggest that although the syllable may function as a perceptual unit for French listeners, for English listeners phonemes are at least as important as syllables, if not more important. But this claim seems to be contradicted by the consistent finding of the monitoring studies reviewed above. English listeners seem to be able to identify syllables more rapidly than phonemes. If the English listeners were *not* using a syllabic segmentation strategy, why should this finding be so robust?

We suggest that faster detection times for syllables than for phonemes are completely artifactual. Replicable as this finding may be, it is due almost entirely to the way in which stimuli have been constructed in most monitoring experiments. Inspection of the materials used in previous comparisons of syllable and phoneme monitoring reveals that in none of the experiments of this type did the nontarget items in the syllable lists ever begin with the same phoneme as the target. Moreover, in most cases none of the remaining phonemes in a given list's target syllable ever appeared in other syllables in a list. As a result, a syllable target could effectively be identified as soon as its initial phoneme had been identified. In all of the existing comparisons of syllable and phoneme monitoring, therefore, it has been possible for subjects to perform the syllable-monitoring task accurately simply on the basis of perception of the initial phoneme of the target. Even Foss and Swinney's (1973) word-monitoring task could be carried out reliably simply by identifying the initial phoneme of the target word. Given this form of list construction, syllable monitoring should always be at least as fast as phoneme monitoring. Subjects are essentially performing phoneme monitoring in both conditions.

On the face of it, this would seem to suggest that syllable- and phoneme-monitoring response times should be indistinguishable. However, it can be argued that the

subjects' task in the syllable-monitoring condition is actually somewhat easier. Subjects performing phoneme monitoring must identify the initial phoneme of the stimulus item. In other words, whenever the stimulus item is more than one phoneme long, the task necessarily involves segmentation of the stimulus items—that is, separation of the phoneme target from adjacent speech. Subjects performing syllable monitoring, however, usually have not needed to segment stimulus items. One reason for this is that many syllable-monitoring experiments have used lists of isolated syllables as stimuli; that is, the targets were bounded by silence rather than by speech. Another reason is that when no other phonemes of the target specification appear in nontarget syllables, subjects may be able to base responses on identification of *any* part of the target item. Thus faster responses in syllable-monitoring conditions may have simply resulted from the relaxation of constraints in these conditions. Subjects could perform their task using only a partial analysis of the stimulus. Subjects asked to detect the syllable *pid*, for instance, may very well have adopted a strategy that would lead them to respond to any syllable containing any of the phonemes /p/, /I/, or /d/.

Any fair comparison of phonemes and syllables should ensure that both phonemes and syllables are analyzed fully. That is, the task should require subjects to be certain that they have distinguished a target phoneme from all other phonemes in the language, and a target syllable from all other syllables in the language. To achieve this, one needs some way of controlling the level of discrimination required in a monitoring task to ensure that subjects cannot respond simply on the basis of a partial analysis of the target. One way to do this is to include in the experiment filler lists in which no item actually matches the specified target, but at least one item very nearly matches it. Such "foil" items should force subjects to adopt a strategy of fully analyzing all items before making a detection response.

In a syllable-monitoring experiment, Mills (1980a) showed that the inclusion of foils that shared the first two phonemes of the target slowed syllable-monitoring latency by almost 150 msec. Of course, the mere presence of foils may itself inflate response times. Therefore, a true test of syllable monitoring versus phoneme monitoring can only be achieved by comparing syllable monitoring in the presence of foils that force complete analysis of each syllable with phoneme monitoring in the presence of foils that force complete analysis of each phoneme.

Complete analysis of the stimulus can be ensured by making foils as similar to the target as possible. For phoneme foils, this can be achieved by having target-foil differences of only one distinctive feature. For example, given the target specification /d/, a list might contain an item beginning with the phoneme /t/, which differs from the target only in the feature of voicing. Similarly, for syllable foils, one phoneme of a stimulus syllable could differ from the target specification by one distinctive feature. For instance, given the target specification *pid*, a list might contain a nontarget item beginning with the syllable *pit*.

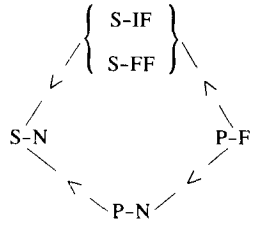
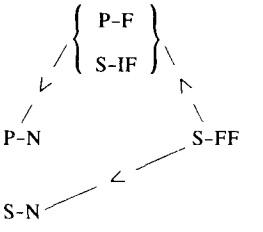
A further form of syllable foil is of interest in testing the strong claim that the syllable is the unit of perception and that phonemic analysis takes place only after the syllable has been identified (Mehler, 1981). Foils (like *pit* after a target *pid*) that are similar to the target in all but the final phoneme can be compared with foils that differ in the first phoneme instead (e.g., *bid* after a target *pid*). If syllables are perceived as units rather than as being constructed from a prior analysis of the individual phonemes, then the foils should be equally similar to targets, whether they diverge at the initial or final phoneme in the syllable. If, on the other hand, listeners are carrying out a phonemic analysis, and syllables are only identified subsequent to completion of this prior analysis, then foils that diverge at the initial phoneme should behave rather like phoneme foils. Only foils that diverge at the final phoneme should function effectively to force a complete *syllabic* analysis.

To determine whether there is actually a response-time advantage for either phonemes or syllables under exactly comparable conditions, we carried out an experiment comparing the different syllable- and phoneme-monitoring conditions described above. That is, we contrasted two phoneme-monitoring conditions, one with and one without foils, and three syllable-monitoring conditions, one with no foils, one with foils that differed from the target on the initial phoneme, and one with foils that differed from the target on the final phoneme.

From strong versions of the perceptual-unit hypothesis, it is possible to derive specific predictions about the ordering of response times in these five conditions. Of course, any theory will predict that the extra analysis forced by the inclusion of foils should lead to an overall increase in latency in conditions with foils as compared with conditions without foils. Therefore, both models agree in predicting that responses in the phoneme-monitoring condition with foils will be longer than responses in the phoneme-monitoring condition without foils, and that responses in the syllable-monitoring conditions with either type of foil will be longer than responses in the syllable-monitoring condition without foils. The hypothesis that the syllable is a perceptual unit, however, claims that foils that diverge from the target at the beginning or at the end of the syllable are equally similar to a syllable target. Therefore, both types of foil should function equivalently, and there should be no response-time difference between the condition with foils diverging syllable-initially and the condition with foils diverging syllable-finally. Additionally, the syllabic-unit hypothesis must predict that syllable monitoring will be faster than phoneme monitoring both in the conditions with foils and in the conditions without foils.

The hypothesis that the phoneme is a perceptual unit, and that phonemic analysis precedes syllabic analysis, makes different predictions. The only true comparison between phonemes and syllables is that between the two con-

Table 1
Predictions of the Syllabic and Phonemic Models

Syllabic Model	Phonemic Model
	
Crucial Predictions	
1. S-FF < P-F	S-FF > P-F
2. S-IF = S-FF	S-IF = P-F

Note—S-N = syllable monitoring, no foils; S-IF = syllable monitoring, syllable-initial foils, S-FF = syllable monitoring, syliable-final foils; P-N = phoneme monitoring, no foils; P-F = phoneme monitoring, foils

ditions that force a full analysis of the target: phoneme monitoring with foils and syllable monitoring with foils that diverge syllable-finally. The phonemic hypothesis predicts that response times will be faster in the phoneme-monitoring condition. Additionally, because only finally-diverging syllable foils will force complete analysis of syllable targets, monitoring in this condition should be slower than monitoring with initially-diverging syllable foils. In fact, since the phonemic hypothesis claims that initially-diverging syllable foils only force analysis of the initial phoneme of a syllable, this hypothesis predicts that such foils will produce response times similar to those of the phoneme-foil condition. Finally, if syllable monitoring without foils is indeed easier than phoneme monitoring without foils, then, in line with previous results, syllable monitoring without foils should be faster than phoneme monitoring without foils.

Table 1 shows the relative ordering of the five conditions as predicted by each model. The crucial predictions that differentiate the two models deal with the relative ordering of the foil conditions. The syllabic model predicts that responses in the syllable-monitoring condition with finally-diverging foils will be faster than responses in the phoneme-monitoring condition with foils, whereas the phonemic model predicts the reverse. The syllabic model predicts that syllable-monitoring responses will be equal with initially-diverging and finally-diverging foils, whereas the phonemic model predicts that responses will be faster with initially-diverging than with finally-diverging syllable foils. In general, the syllabic model predicts that the slowest condition overall will be phoneme monitoring with foils, while the phonemic model predicts that the slowest condition will be syllable monitoring with finally-diverging foils.

METHOD

Experimental Design

Two phoneme-monitoring conditions, one with and one without foils, were contrasted with three syllable-monitoring conditions, one with no foils, one with foils that differed from the target on the initial phoneme, and one with foils that differed from the target on the final phoneme.

Subjects in all conditions heard exactly the same sequences of stimulus items. Within the phoneme-monitoring conditions, all subjects responded to the same phoneme targets. The only difference between the conditions was in the target specification for a subset of 20 filler trials on which no response was appropriate: in the foil condition, these filler trials had target specifications that differed by one distinctive feature from the initial phoneme of some item in the sequence, whereas in the nonfoil condition, these trials had target specifications unlike any initial phoneme in the sequence. Similarly, in the three syllable-monitoring conditions, all subjects responded to the same syllable targets (which were the initial syllables of the same items to which the phoneme-monitoring subjects responded). The only difference between the three conditions was in the same subset of no-response filler trials. In one syllable-monitoring condition, these trials had target specifications that were unlike any initial syllable in the sequence. In another, the target specification for these trials differed from the initial syllable of some item in the sequence by one distinctive feature of the initial phoneme. In the third condition, the target specification for these trials differed from the initial syllable of some item in the sequence by one distinctive feature of the final phoneme.

As an example, one response trial was "pastry spartan pilot gamble hot." The phoneme target was /g/ and both phoneme-monitoring conditions received this target specification. The syllable target was *gam* and all three syllable-monitoring conditions received this target specification. One of the crucial no-response filler trials was "ashes guest willow harmony fattening orange." The phoneme-monitoring no-foil condition (P-N) received the target specification /p/ for this sequence, and the phoneme-monitoring foil condition (P-F) received the target specification /v/, which differs from the initial phoneme of "fattening" by only the feature of voicing. For the same sequence, the syllable-monitoring no-foil condition (S-N) received the target specification *pem*, the syllable-monitoring initial-foil condition (S-IF) received the target specification *vat*, and the syllable-monitoring final-foil condition (S-FF) received the target specification *fad*.

In summary, the five conditions in the experiment differed only in the target specifications that were presented to subjects. All subjects listened to a single identical set of auditory stimuli. Subjects in two conditions performed phoneme monitoring; their specified targets were phonemes. The specifications for the two conditions differed only in the 20 foil sequences; the specifications for the experimental sequences were identical for both conditions. Subjects in the other three conditions performed syllable monitoring, and hence were presented with syllable targets. Again, the target specifications for the experimental sequences were identical in all three conditions; the conditions differed only in respect to the target specifications for the 20 foil sequences.

Materials

The target items in the experiment were a set of 20 polysyllabic words and 20 nonwords. By using polysyllabic items, we avoided having the level of items in the sequence match the level of one of our target types and not match the other (Healy & Cutting, 1976; McNeill & Lindig, 1973); polysyllabic items were a mismatch both to phonemes and to syllables. By using nonwords, we were able greatly to increase the size of our materials sets; the word sequences were so highly constrained by our requirement of keeping the complete set of targets constant across conditions that a larger set would have been difficult to achieve. A further 185 words and 185 nonwords were chosen, varying in length from one to three syllables. These were used to fill out the sequences in which the target items appeared, and to construct the practice and foil sequences. The complete set of items is shown in the appendix. Each target word was matched with a target nonword on number of syllables and initial

phoneme, and on position in which each occurred in its respective sequence. Two tapes were created, one containing only words and one containing only nonwords. Each tape consisted of 10 practice sequences and 35 experimental sequences. The sequences were between one and six items in length, and the experimental targets appeared in the second, third, fourth, or fifth position. The experimental target was always the penultimate item in the sequence. There were also five filler targets that occurred in the first position in a sequence, two in the second position, and one in the sixth position.

There were 10 no-response sequences per tape, 3 in the practice set and 7 in the experimental set. These sequences were those on which foil targets occurred in the foil conditions. The sequences themselves were identical across all five experimental conditions. Different conditions were created by altering the target specification that was presented visually before the start of each sequence. In the S-N conditions, both the initial and final phonemes in the target specification differed from the corresponding phonemes in all of the initial syllables in the sequence by at least two distinctive features. For example, they could differ both in voicing and in place of articulation. In the foil conditions, too, all items except one differed from the target specification by at least this amount. The target specification for a foil sequence shared two phonemes with the initial syllable of some item in the sequence, differing from the initial syllable of that item by only a single distinctive feature of a third phoneme. In the S-FF condition, the final phoneme differed, and in the S-IF condition, the initial phoneme differed. The target specifications for the P-N and P-F conditions were always the first letters of the S-N and S-IF conditions, respectively. Within both the phoneme and syllable conditions, the different foil conditions were created by rearranging the assignment of target specifications to sequences so that all subjects performing syllable monitoring saw the same set of syllable targets, and all subjects performing phoneme monitoring saw the same set of phoneme targets.

All subjects heard both the word and the nonword tapes. Half of the subjects in each condition heard the word tape first, and the other half heard the nonword tape first.

To eliminate the possibility of carry-over effects from foil to no-foil conditions, it was essential to run the different foil conditions as independent groups. Unfortunately, between-subjects designs tend to lack sensitivity because group differences are often swamped by between-subjects variance. We therefore adopted a measure designed to increase the sensitivity of the experiment by assessing each subject's overall speed in an auditory monitoring task and analyzing the results as a covariate of the results in the main monitoring task. For our covariate measure, we required a task that would be as similar as possible to syllable and phoneme monitoring while being sufficiently different that there would be no carry-over effects between the covariate task itself and the main monitoring task.

The task chosen was an auditory monitoring task using nonspeech stimuli. The subjects were required to listen to sequences of between two and six tones presented over headphones. In each sequence, the target tone was a square wave with a mark-to-space ratio of 2:1. The remaining tones had a mark-to-space ratio of 1:1. The tones varied in frequency from approximately 75 Hz to approximately 125 Hz. There were 30 sequences of tones, 20 of which contained experimental targets. There were seven trials without targets, and responses to the remaining three trials with targets (the first three in the experimental set) were not recorded. The 30 experimental sequences were preceded by 10 similar practice sequences. The tone-monitoring covariate task was always presented before the syllable- or phoneme-monitoring task and was presented as a completely separate experiment. At the beginning of the session, the subjects were given examples of both target and nontarget tones and were instructed that their task was to press the response button as quickly as possible as soon as they heard a target tone.

After all subjects had been tested, a minor error was discovered in the target-specification lists. Two syllable target specifications in the word list had been inadvertently transposed (a practice item that should have had the target *bam* was given the target *bat*, and an experimental item near the end of the word list that should have had the target *bat* received the target *bam*). The error did not affect the phoneme-monitoring conditions, nor did it affect the assignment of target specifications to foil conditions. Also, the final sounds differed by more than one distinctive feature, so that the syllables involved were not as alike as those in the S-FF condition. Nevertheless, the effect of the error was that subjects in both the S-N condition and the S-IF condition received two trials on which the target specification differed from the initial syllable of some item in the list by only the final consonant, whereas only subjects in the S-FF condition should have received such trials. Of course, any effects of this error would be equally opposed to both the phonemic hypothesis and the syllabic hypothesis because the differences predicted between these three conditions would be reduced. We decided (1) to remove responses to that experimental item from the syllable-monitoring conditions (in fact, this should not have been necessary; no subject should have responded to that item, since it did not in fact match the target), and (2) to carry out an analysis of responses to the nonword list as a function of order of list presentation. Since the error was in the word list, responses to nonwords by the subjects who had the nonword lists first should be unaffected by it, whereas responses to nonwords by the subjects who heard the nonword list after the word list should be susceptible to any effect the error might have.

Subjects

The subjects were 138 members of the Applied Psychology Unit panel of volunteer subjects recruited from the Cambridge community. The age range was 19 to 49 years (mean age: 33). The subjects were paid a small fee for participating in the experiment.

Procedure

The subjects were tested individually in a sound-attenuated room. The subjects wore headphones and were seated in front of a video display unit (VDU) controlled by a microcomputer. To ensure that the subjects paid attention to the target specification in the speech-monitoring task, the VDU bell was sounded as each specification appeared on the screen. Response times were measured from an inaudible tone placed at the onset of each target item. In order to check whether subjects were erroneously responding to the foil items, response times to these items were recorded in the same manner. To simplify the running of the experiment, the word and nonword tapes were spliced together. All subjects in the word-nonword order were run first, after which the tapes were respliced so that the remaining subjects heard the nonword condition before the word condition. The subjects were assigned to the five foil conditions in the order in which they arrived for the experiment.

If the subjects were performing their task correctly, they should have responded to all of the experimental targets but none of the foil targets. Therefore, we recorded responses on the 14 no-response trials in the experimental sets, and any subjects who responded on more than 4 of these were replaced. Excluding those who were rejected for this reason, 24 subjects were tested in each of the five conditions, with tape order counterbalanced within conditions.

RESULTS

The mean reaction times, both raw and adjusted for the covariate, are shown in Table 2. The data are in line with the predictions of the phonemic model and opposed to the

Table 2
Mean Raw and Adjusted Response Times (Msec)

	P-N	P-F	S-N	S-IF	S-FF
	Unadjusted Mean Reaction Times				
Words	464	529	526	541	655
Nonwords	477	508	461	494	564
Mean	470	519	494	518	610
Covariate	380	401	386	366	366
	Means Adjusted for Covariate				
Words	464	511	519	551	665
Nonwords	478	487	455	508	579
Mean	470	499	487	530	622

Note—P-N = Phoneme monitoring, no foils; P-F = phoneme monitoring, foils; S-N = syllable monitoring, no foils; S-IF = syllable monitoring, syllable-initial foils; S-FF = syllable monitoring, syllable-final foils.

predictions of the syllabic model. The slowest condition of all was S-FF, exactly as predicted by the phonemic model. Indeed, the overall results appear to support a rather stronger version of the phonemic model than that described earlier, in that phonemes are identified slightly faster than syllables even in the no-foil conditions.

An analysis of variance was first conducted on the raw response times. The main effect of groups was highly significant [$F_1(4,110) = 5.24, p < .001$]. Targets on nonwords were detected faster than targets on words [$F_1(1,110) = 46.3, p < .001$]. There was also an interaction between these two effects [$F_1(4,110) = 8.36$, $p < .001$], which was due to subjects in the P-N group producing faster response times to targets on words than to targets on nonwords, in contrast to the other four groups. The main effect of tape order was not significant, and did not interact with either of the other variables.

Planned comparisons were carried out on the means adjusted for the covariate measure. Recall that the most important comparisons for distinguishing between the two models were those between the various foil conditions. In each case, the phonemic model's predictions were supported and the syllabic model's were not. As predicted by the phonemic model, responses in the P-F condition were significantly faster than those in the S-FF condition [$t(109) = 4.81$, words $t(109) = 5.48$, nonwords $t(109) = 3.47$]. Again, as predicted by the phonemic model, syllable-initial foils led to faster responses than did syllable-final foils [$t(109) = 3.6$, words $t(109) = 4.06$, nonwords $t(109) = 2.68$]. Responses in the S-IF condition and the P-F condition were not significantly different [$t(109) = 1.21$, words $t(109) = 1.42$, nonwords $t(109) = .79$]. These comparisons conclusively make the case in favor of the phonemic hypothesis. When full analysis of the target is compulsory, phonemes are detected faster than syllables. And syllables are not processed as unanalyzed wholes, because syllable-initial foils and syllable-final foils are not equally effective at forcing full analysis of the syllable. The difference between the syllable-initial and syllable-final foil conditions also indicates that the slow responses in the S-FF condition were not simply due to the presence of foils as such.

The remaining comparisons involved the conditions without foils. As predicted by both models, syllable monitoring with no foils was faster than with either finally-diverging or initially-diverging foils [S-N vs. S-FF: $t(109) = 5.28$, words $t(109) = 5.2$, nonwords $t(109) = 4.68$; S-N vs. S-IF: $t(109) = 1.68$, words $t(109) = 1.14$, nonwords $t(109) = 2.0$; only the comparison in the words condition failed to reach significance at the .05 level on a one-tailed test]. However, phoneme monitoring without foils was not significantly faster than that with foils [$t(109) = 1.11$, words $t(109) = 1.67$, nonwords $t(109) = .34$; the comparison in the words condition just reached significance at the .05 level on a one-tailed test]. The final comparison dealt with the relationship between syllable and phoneme monitoring in the absence of foils—the comparison supposedly made in previous tests of syllable versus phoneme monitoring. The syllabic hypothesis clearly predicts that S-N should produce faster responses than P-N. However, the present results failed to support this prediction. In fact, the response-time difference was actually in the opposite direction, although it reached significance only in the words condition [$t(109) = 0.65$, words $t(109) = 1.96$, nonwords $t(109) = .87$].

Simple between-condition t tests on the raw means exactly mimicked the pattern of the planned comparisons on the means adjusted for the covariate.

In order to determine whether our error in the target-specification list had affected response times, we first inspected responses to the one affected experimental item. As expected, no subjects in the S-FF condition had responded to this item, whereas 3 of 24 subjects in the S-IF condition and 4 of 24 subjects in the S-N condition had erroneously responded. These seven responses were discarded. An analysis of variance was then conducted on the nonword responses as a function of order of presentation of the lists. There was no effect of order either as a main effect ($p > .1$) or, more importantly, as an interaction with the condition means ($p > .8$). It was con-

cluded that the error had had no significant effect on responses.

DISCUSSION

Phoneme-monitoring response times are faster than syllable-monitoring response times. The single most significant result of this experiment is that responses in the P-F condition were very much faster than those in the S-FF condition. Only in these conditions can one be sure that the subjects fully analyzed the targets before responding.

The effect of the foils in the S-FF condition can be seen even more clearly if we compare the subjects who made fewer than five errors on the foil trials with the subjects who were rejected because they exceeded this error criterion. In the S-FF condition, 18 subjects had to be rejected on the basis of their errors to the foils in order to get 24 subjects who passed the error criterion. In neither of the other foil conditions did we have to reject any subjects at all. Clearly, the subjects found it very difficult to avoid responding before they had analyzed the whole syllable. Additional evidence that subjects who made excessive errors to the foils were responding prematurely comes from an examination of their overall reaction times. In a further analysis of covariance involving all subjects in the S-FF condition, subjects who made five or more errors were found to have responded 112 msec faster than those who made fewer than five errors (see Table 2). However, this difference was only marginally significant [$t(36) = 1.79$, $0.1 > p > .05$]. It seems that speed in the syllable-monitoring task can only be increased by responding before the syllable has ended, a strategy that led to an increase in errors on the S-FF trials.

If the syllable really were the unit of perception, subjects would have no choice but to process the entire syllable before responding. Therefore, this evidence of premature responses to syllable targets provides a clear indication that our listeners were *not* processing the input syllable by syllable, but instead were analyzing it in a left-to-right fashion at a level below the syllable. Our results are thus perfectly in accord with a considerable body of recent evidence favoring left-to-right phonemic or phonetic processing in the perception of English (e.g., Cole & Jakimik, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Welsh, 1978; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985; Warren & Marslen-Wilson, 1987).

Two aspects of our data call for further comment. First, the differences we found were, with one exception, larger in the case of words than in the case of nonwords. It seems doubtful that great importance should be attached to this. The level of significance was in most cases identical for words and for nonwords, and it will be recalled that responses to nonwords were overall significantly faster than to words; this suggests that the differences could have been attenuated via a simple floor effect. Moreover, there may have been structural differences between our word and nonword stimuli that could have produced such a difference. In an attempt to maximize the potential value of a syllabically based analysis, we tried to choose words that had clear syllable boundaries. Given the prevalence of ambisyllabicity in English, however, it was difficult to select words that conformed to this as well as to all our other requirements. In the nonword conditions, though, we were able to construct purpose-built stimuli. These nonwords may have therefore had clearer syllable boundaries than the words, which would have given syllable monitoring an advantage in nonwords in our experiment, even though such an advantage would not be characteristic of English words in general.

Second, an aspect of the present data that was not predicted by the phonemic hypothesis was the finding that phonemes were identified rather faster than syllables even in the no-foil conditions. On the basis of previous results in the literature, we would have expected syllables to have been identified significantly faster than phonemes under these conditions. However, the present experiment failed to replicate this robust result. We now believe that the faster detection times previously reported for syllables than for phonemes have been entirely artifactual. Although the failure to include foil items has been the chief factor in this spurious finding, there have also been other aspects of the design of previous studies that, we believe, have assisted in producing a response-time advantage for syllable over phoneme targets.

For example, in the present experiment, all of the target-bearing items were polysyllabic words or nonsense items. This was also true of the materials used by, for instance, Foss and Swinney (1973) and Segui et al. (1981). In contrast, other experiments used sequences of syllables (e.g., Savin & Bever, 1970; Swinney & Prather, 1980). McNeill and Lindig (1973) showed that the use of syllable sequences will tend to produce faster responses to syllables because syllable targets match the level of all items in a sequence. Moreover, it is also the case that if syllable targets appear in sequences of syllables, then there is no need to segment the input in order to isolate the syllable and match the input against the target specification. In such experiments, therefore, syllable monitoring would be comparatively easy. Phoneme monitoring, on the other hand, would be somewhat harder because the target-bearing syllable must be segmented before a successful match can be achieved (see Norris & Cutler, 1985, for a discussion of the relation between segmentation and identification). Phonemes will therefore tend to be responded to more slowly than syllables.

Healy and Cutting (1976) also used sequences of syllables, but their phoneme targets were vowels that could appear either as part of a VC syllable or in isolation. Of course, given that vowels in isolation are syllables, one could argue, as we pointed out above, that Healy and Cutting simply compared syllable monitoring with syllable monitoring. Using these conditons, Healy and Cutting failed to replicate the response-time advantage for syllables over phonemes.

Finally, Foss and Swinney (1973) also used bisyllabic words, this time, of course, in English. Foss and Swinney had subjects monitor for phonemes, syllables, or

words, with the level of the target changing from trial to trial. Word monitoring produces a shift of attention toward the word level and away from the phoneme level, in comparison to phoneme monitoring (Brunner & Pisoni, 1982). Any such shift is likely to benefit syllable monitoring at the expense of phoneme monitoring. Many of Foss and Swinney's syllable targets were in fact meaningful words in themselves. Foss and Swinney compared meaningful with meaningless syllable targets and found that meaningful targets were responded to somewhat more rapidly.

However, a detailed examination of Foss and Swinney's (1973) stimuli provides an even more satisfying explanation for why their results differed from ours. Foss and Swinney kindly provided us with a complete list of their materials, which revealed that their syllable-monitoring condition contained no foils and was therefore similar to our no-foil condition. Their phoneme-monitoring condition, however, contained a far higher proportion of foils than even our phoneme-foil condition. Overall, 47% of their stimulus lists contained some kind of phoneme foil. Of their 11 no-response phoneme-target filler lists, 5 had /b/ as target and contained a word beginning with /p/; a further 2 had /b/ as target and contained words beginning with /d/. The other 4 lists all contained an occurrence of the specified target within some word in the list. In addition to these filler foils, there were a large number of foils preceding the targets in the 51 experimental lists (of which 17 were phoneme-target lists for each of their three subject groups). Ten of these lists contained, prior to the occurrence of the target-bearing item, a word beginning with a phoneme foil (i.e., a phoneme differing from the target phoneme by only a single feature); 17 had the target phoneme itself appearing somewhere within a word in the list; and a further 9 had both of these features! Of the 100 lists that were used in Foss and Swinney's experiment, 67 (51 plus 16) occurred with a phoneme target; of these, 47 (70%) contained some form of phoneme foil.

In contrast, 68 lists (51 plus 17) occurred with a syllable target; none contained a syllable foil. Sixty-seven lists (51 plus 16) occurred with a word target; none contained a word foil.

Given the construction of these lists, it is not at all surprising that Foss and Swinney (1973) found syllable monitoring to be speedier than phoneme monitoring. Their experiment clearly involved a comparison of a phoneme-foil condition with a syllable-no-foil condition. Their results are therefore fully consistent with our own finding that syllable monitoring in the syllable-no-foil condition is faster than phoneme monitoring in the phoneme-foil condition.

The fact that Foss and Swinney's (1973) materials contained phoneme foils in the experimental lists as well as in the filler lists raises yet a further problem with their experiment. Nineteen of their 51 experimental lists contained a word beginning with a phoneme foil appearing before the target item. Newman and Dell (1978) showed that phonological similarity between the target and the initial phoneme of preceding words can inflate reaction times by over 200 msec. Therefore, quite independent of any effect of the phoneme foils in filler trials, the effect of foils in experimental trials is probably sufficient to account for Foss and Swinney's finding that phoneme monitoring is slower than syllable monitoring.

We would argue, then, that no previous experiment has properly compared detection of syllable targets with detection of phoneme targets. There have been many experimental design features that have biased the results of previous experiments in favor of faster responses to syllables. By far the major problem, though, has been the failure to include foil items. This has allowed subjects to get by with partial analysis of the stimuli in syllable-monitoring conditions. That is, the advantage of syllables over phonemes observed in earlier experiments is principally due to the fact that subjects have been able to perform syllable monitoring simply by identifying the initial phoneme of the target. As the present results clearly demonstrate, however, when subjects are forced to analyze syllables and phonemes fully, phonemes can be identified faster than syllables.

Our results, in conclusion, undermine claims that the syllable is the major unit of perception—at least in English. However, it should be pointed out that these results do not allow us to propose an alternative perceptual unit valid for all languages. First, there is no guarantee that the present results would be replicable in other languages. As we pointed out in the introduction, there is evidence that French listeners, for example, syllabify speech input in a way English listeners do not. Perhaps, therefore, French listeners would perform differently in the present monitoring tasks—particularly with respect to the crucial comparison between the effects of initially-diverging and finally-diverging foils. We hope that the present study may be replicated in one or another of the languages that, unlike English and other stress-timed languages, appear to lend themselves to syllabification as a segmentation strategy. Second, we would stress that a comparison between monitoring tasks, such as we have performed, cannot be interpreted as directly addressing the issue of levels of representation in speech understanding. Successful detection of targets is not dependent on the existence, in normal speech understanding, of a level of representation corresponding to the level of the target. Although we have shown that phonemes can be identified faster than syllables in English, this should not be taken as direct evidence that phonemes are the unit of perception in English. What we have shown is that speech can be analyzed in a left-to-right manner at some level below the syllable. This level could be the phoneme; but it could equally well be an acoustic template from which a phonemic representation on which to base the monitoring response can be derived. Whatever the level of the initial perceptual analysis, though, it is clear that, at least in English, a phonemic representation can be derived before a syllabic one.

REFERENCES

BRUNNER, H., & PISONI, D B (1982). Some effects of perceptual load on spoken text comprehension. *Journal of Verbal Learning & Verbal Behavior*, **21**, 186-195.

COLE, R. A., & JAKIMIK, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America*, **67**, 965-970.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1983). A language specific comprehension strategy. *Nature*, **304**, 159-160.

CUTLER, A., MEHLER, J., NORRIS, D., & SEGUI, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory & Language*, **25**, 385-400.

FOSS, D. J , & SWINNEY, D. A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *Journal of Verbal Learning & Verbal Behavior*, **12**, 246-257.

HEALY, A., & CUTTING, J. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning & Verbal Behavior*, **15**, 73-83.

MARSLEN-WILSON, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention & Performance X* (pp. 125-150). Hillsdale, NJ: Erlbaum.

MARSLEN-WILSON, W. D., & WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29-63.

MCNEILL, D., & LINDIG, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning & Verbal Behavior*, **12**, 431-461.

MEHLER, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, Series B*, **295**, 333-352.

MEHLER, J., DOMMERGUES, J.-Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning & Verbal Behavior*, **20**, 298-305.

MILLS, C. B. (1980a). Effects of context on reaction time to phonemes. *Journal of Verbal Learning & Verbal Behavior*, **19**, 75-83.

MILLS, C. B. (1980b). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 528-535.

NEWMAN, J., & DELL, G. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning & Verbal Behavior*, **17**, 359-374

NORRIS, D., & CUTLER, A. (1985). Juncture detection. *Linguistics*, **23**, 689-705.

PISONI, D. B., NUSBAUM, H. C., LUCE, P. A., & SLOWIACZEK, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, **4**, 75-95.

SAVIN, H. B., & BEVER, T. G. (1970). The non-perceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, **9**, 295-302.

SEGUI, J., FRAUENFELDER, U., & MEHLER, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, **72**, 471-477.

SWINNEY, D., & PRATHER, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, **27**, 104-110.

WARREN, P., & MARSLEN-WILSON, W. D. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, **41**, 262-275.

APPENDIX
Experimental Materials

Foil lists are labeled "f." Their targets for the five groups are listed in the order P-F, P-N, S-FF, S-IF, S-N.

Phoneme targets for nonfoil lists were always the first phoneme of the syllable targets. Reaction times in Table 2 are for the 40 experimental lists only (numbers italicized).

Words: Practice Set

1. MED house feather video weed estuary *medical*
2. FAM hazard mountain *family* warmer
3. f cheaper rage myth *gantry* sailor
   C,T,GAM,CAN,TAV
4. PAD garbage *padlock* leather
5. DEC lettuce post *decorate* box
6. CAN dale wings jailer feet delight *candid*
7. f rust fortune jest *nibble* pastry
   M,C,NIP,MIB,COC
8. VIC *victory* damage
9. f lifting hopeless ruler wish *tavern* laugh
   D,B,TAF,DAV,BAB
10. BAM cattle file crash *bamboo* erase

Words: Experimental Set

11. GOB rapid *goblin* trip
12. CON days forest slave *convent* earnings
13. PAN figure charming *pancake* marrow
14. f evolve ration sharpen list *pencil* foil
    B,N,PEM,BEN,NIP
15. BAN alchemy *banter* lavender
16. DET shilling luggage *detonate* arrest
17. MAG carpet abbot ladder bottle *magnitude* old
18. COB *cobweb* lawn
19. f arrangement each lather *mammal* editor
    N,V,MAN,NAM,VEGG
20. VAM educate pastry racing caress *vampire* grain
21. VIN personal sparrow pardon dirt *vindicate* rapid
22. BAF staple audition *baffle* vet
23. f laugh vole *definite* rascal
    T,M,DEV,TEF,MAN
24. NAT *natural* shampoo
25. COM harvest fibre reject flatter *combat* leaf
26. NEG petrol design *negligent* bat
27. PEN seldom goal tide dove *pendulum* devil
28. f nervous risk list *baptist* shield
    P,F,BAB,PAP,FAD
29. GON include cellar *gondola* radio
30. VAC grass telephone shadow hobble male *vacuous*
31. FAB callous sound lemon shallow *fabricate* gale
32. f essay lunatic holiday *cognizant* marriage
    G,D,COC,GOG,DEV
33. NAP armour craving wave charm *napkin* carton
34. GAM pastry tartan pilot *gamble* hot
35. TAN vague sable bolster food *tantrum* guess
36. MAT *matter* relay
37. f ashes guest willow harmony *fattening* orange
    V,P,FAD,VAT,PEM
38. BAT cable soap harmful *battle* condone
39. TAB remark milk *tablet* boiler
40. PUD *puddle* steam
41. MAN brain spartan *mandate* guard
42. f pelt salad storage cheese *vector* harp
    F,G,VEGG,FEC,GAM
43. DEC sensible famous short *decrement* slip
44. FAC cabbage poet *factory* handle
45. TEN *tentative* pale

Nonwords: Practice Set

46. KEP albit belig vade *kepsin* hoffe
47. f kaldat shaste leel thutch losh *nebdim*
    M,F,NEP,MEB,FIC
48. FOD *fodrage* manel

49. f asbensing chidder forn *tanlist* jesh
D,G,TAM,DAN,GOB
50. PAG wennit murrows falt ood *paggim* nosk
51. MAL kaffik greble *malate* cheg
52. DEN faffle *dennelled* eld
53. f chole vone fesh *goplam* halaram
C,M,GOB,COP,MED
54. BED ranthin hivin *bedrel* nojis
55. VIN beshful sedum masik drilazik treeler *viniple*

Nonwords: Experimental Set

56. MOV pendle *movander* anding
57. FID brenning gimit *fiddeny* conderate
58. DAP kem thastin chent *dapmatiss* freg
59. f helst sparth wilth choffe *vidma* alshim
F,T,VIT,FID,TAM
60. MAR kurabad fash *marbate* watters
61. GUD *guddle* foon
62. TIB mordage alse *tiblem* evel
63. BEM cravel threck *bemday* keedle
64. f oshel chivish edding nabler *figstan* thetchin
V,B,FIC,VIG,BAV
65. DAM *damik* kosin
66. TAD balvish stope shalun fook *tadrum* vone
67. GOM slape plote javon *gomble* thig
68. NAM keendis elvint scamming trowik *namsin* lorim
69. f monic lemis septil *kovlish* holimul
G,D,COF,GOV,DAG
70. FEN cadalid madle gabet thale *fenilate* bettish
71. PEC grib stram jeest foril eedat *peckist*
72. GAF chander spet pannicle *gafted* jutton
73. f shardis throdle lalt *baftim* chid
P,N,BAV,PAF,NEP
74. PEG zurble gazil freck losk *peggilum* hild
75. NUG kaab credole *nugsarent* jesk
76. KED shoning jellip adjed pont *kedvet* fam
77. NAT *natrum* ellant
78. f savish yage *dakfar* raich
T,P,DAG,TAC,PEN
79. BAV krod rellin *bavray* chesht
80. VID joller kavaling grisht jal *vidlikish* sim
81. VOG densiv hardle derson thropper *voglor* dat
82. f debling levid arging *metstrin* losh
N,C,MED,NET,COF
83. TAF *taffic* lumard
84. MEG thope shapple horsin jestin *megdoh* felk
85. DEP bekate errak *depanate* halb
86. CAV seltish *kavish* heffible
87 f lelt gam hoge shisle *pemlin* choft
B,V,PEN,BEM,VIT
88 PIN dack lalin *pinmape* holin
89. CAM ediding cheg alsh *kampent* thesh
90. BAN chelt *bandil* fost

(Manuscript received March 11, 1987;
revision accepted for publication November 10, 1987.)