JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **12**, 246-257 (1973)

On the Psychological Reality of the Phoneme: Perception, Identification, and Consciousness[1]

DONALD J. FOSS AND DAVID A. SWINNEY

University of Texas at Austin, Austin, Texas 78712

In Experiment I, 40 subjects listened to 100 lists of two-syllable words. They monitored each list for the presence of either a specified word or an initial syllable of a word. Reaction times to respond to these targets were recorded. Reaction times were significantly shorter when the target was a two-syllable word. In Experiment II, 45 subjects listened to the lists and monitored for either a two-syllable word, the initial syllable of a word, or the initial phoneme of a word. Reaction times were shortest for word targets and longest for phoneme targets. Implications of these findings for the "reality" of the units was discussed. A distinction between perception and identification was introduced and, in addition, a proposal about a determinant of consciousness was forwarded.

One of the major problems to be faced by any model of sentence comprehension results from the fact that the acoustic signal is not reliably segmented into "language-sized" units. The work of Liberman and his associates (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) has amply demonstrated that there is rarely a one-to-one correspondence between the acoustic signal and the units of perception. The same signal will be perceived differently in different linguistic environments and, in some cases, quite different signals will be perceived as being identical. In addition, all the important acoustic cues for some perceptual entity may not occur at the seemingly proper or expected place in the speech stream. In some cases, for example, a critical cue for a consonant is given by the relative length of the preceding vowel (Chomsky & Miller, 1963). These facts indicate that the speech decoding processes cannot be simple ones. The present paper is primarily concerned with characterizing the units of speech perception. This concern eventually leads us to a more general discussion of speech perception.

In 1970, Savin and Bever reported some experimental results that have important implications for any characterization of speech perception units. In one of their experiments (Experiment II) Savin and Bever presented subjects with lists of nonsense syllables, as in (1), spoken at the rate of 600 msec/syllable.

(1) thowj tuwp sarg tiyf friyn ...

On one-half of the lists the subjects were told to respond by pressing a button as soon as they heard a syllable that began with /s/. This is called a phoneme-monitoring condition. (It should be noted that /s/ is one of the few phonemes that does have a set of reliable, context-independent cues in the speech stream. This is further discussed below.) On the other lists the subjects were told the entire target syllable before the list was presented (e.g., "The target is *sarg*"). This is called a syllable-monitoring condition. Reaction times (RTs) for phoneme- and syllable-monitoring were collected and compared.

Savin and Bever's results were unequivocal

[1] These studies were supported by NSF Grant GS-3285 to the University of Texas at Austin. The authors thank B. Pritchett for aid in gathering the data in Experiment II and K. Gummerman and D. Hakes for valuable criticisms of an earlier version of the paper.

and, perhaps at first glance, somewhat surprising. The phoneme-monitoring RTs were significantly longer (354 msec) than the syllable-monitoring RTs (311 msec). Although in the case of /s/, information about the identity of the phoneme occurs earlier than does the information about the identity of the syllable, the RTs certainly did not reflect this fact. Since the subjects were slower in responding to phonemes than to syllables, Savin and Bever concluded (p. 300), "We take our results to show that phonemes are perceived only by an analysis of already perceived syllables (or at least already perceived consonant-vowel pairs)." They went on to argue that while phonemes are psychologically real constructs, they are not perceptual entities but, rather, are "abstract."

While the above was the major outcome of the Savin and Bever work, two other findings are also of interest. First, in another experiment the target was the stop consonant /b/ or syllables which began with this phoneme. The RTs to monitor for these targets were generally shorter than those observed when the target was /s/ or syllables that began with /s/. This is notable since the stop consonants are much more highly encoded in the acoustic signal than is the phoneme /s/. More of the surrounding signal needs to be examined in the former case before the phoneme can be perceived. Nonetheless, the subjects were faster in responding to the more highly encoded target. The second notable result obtained by Savin and Bever was that the RTs were very short relative to the duration of the stimuli. The syllables lasted 600 msec while the responses to them generally were completed in under 300 msec. The subjects were obviously not waiting to hear the entire syllable- before initiating their responses.

Each of the abovementioned results leads to a number of further questions. In addition, there are two other considerations that taken together strongly suggest that these results be further investigated. First, Savin and Bever's inference that phonemes are perceived only by an analysis of already perceived syllables implies that the perceiver gains access to higher linguistic units, such as morphemes, via a coding system that does not take into account the internal structure of the syllable. This means that the memory representation of morphemes or other lexical items would he in terms of a syllabary since the syllable would be the unit of lexical access. Savin and Bever state that once the syllable is recognized, then information about its components can be recovered. Perhaps such information is stored with the syllable in much the same way that syntactic and semantic information is stored with the morphemes. But, to repeat, since the primary access code is via the syllable, the major storage code must then be a syllabary.

It is notable that the syllables of interest here would have to be phonetic syllables. That is, the syllable [p^haet^h] with a released final consonant would be different from the syllable [p^haet] without the final release. Since the perceptual system is blind to the internal structure, of these syllables, one could as well be represented as $Sy1_{239}$, the other as Syl_{240}. At this level of analysis, then, [p^haet^h] and [p^haet] are as different from one another as are [p_haet] and [p^haed]. Hence, *each* will have to be represented at the higher level of analysis along with its (identical) associated syntactic information. Thus, the representation of the input in terms of phonetic syllables will be very uneconomical. While we cannot say a priori that the resulting ineconomy is psychologically incorrect, it does seem to be undesirable if there is an alternative. This problem can be elaborated (see Foss, in press, for some further discussion), but enough has been said for our present purpose, which is to indicate that representing phonological information in the lexicon primarily at the level of the phonetic syllable is at least questionable.

The second consideration to be briefly discussed concerns the view expressed by Savin and Bever that phonemes are "abstract" and, by inference, that syllables are "real." This view naturally leads one to ask for the criterion

that determines the reality status of syllables. The difficulty of providing a physically based criterion for dividing the speech signal into syllables is well known (see Kozevnikov &, Cistovic, 1965). To some considerable extent the syllable is abstract too. To quote Jakobson and Halle (1968, p. 418), "According to opinion ... latently tinging the writings of various authors, phonemes are abstractional, fictitious units. As long as this means nothing more than that any scientific concept is a fictional construct, such a philosophical attitude cannot affect phonemic analysis. Phoneme, in this case, is a fiction in the same way as morpheme, word, sentence, language, etc." It should be quickly pointed out that Savin and Bever do not consider the phoneme to be a mere fiction, a data-summarizing term. They explicitly state that they consider the phoneme to be "real" in the sense that it is a construct required for correct linguistic and psycholinguistic analyses. The only point to be made here is that the line between the phoneme and the syllable that they seem to want to draw, a line separating the abstract from the nonabstract, is not at all clear.

Considerations such as those discussed above motivated.a further exploration of the Savin and Bever results.

EXPERIMENT I

The basic idea behind Experiment I was this: If one concludes that the syllable is perceptually "real" and the pheneme "abstract" from the finding that RTs to syllables are shorter, one would likewise be forced to conclude that syllables are "abstract" if RTs to two-syllable words were shorter than to the syllables beginning those words. If this conclusion seems unwarranted and unpalatable (and it does, for reasons further discussed below), then the initial conclusion is also unwarranted.

In this study the subjects were presented lists of two-syllable words, as in list (2):

(2) rural ladder import candy woven ....

They were asked, to monitor for either an entire two-syllable word (e.g., the word *candy*) or for the first syllable of the word (e.g. the syllable *can*). The RTs to respond to the target were measured and compared.

In addition, the nature of the single-syllable targets was manipulated. On one-half of the lists the single syllable itself constituted a meaningful word (e.g., the syllable *can* in the above example). On the other lists the first syllable did not constitute a meaningful item when taken alone. If the status of the first syllable, meaningful or not, has an impact on the RTs, then it would seem that some rather high-level processing must occur before the response is initiated. This will be further discussed below.

Method

Subjects. The subjects were 40 undergraduate psychology students serving in partial fulfillment of a course requirement at the University of Texas at Austin.

Materials. The materials consisted of 100 lists of two-syllable English words. The lists were from five to nine words long with equal numbers of lists at each length. Fifty of the lists contained target words and 50 did not. The latter were included to prevent increased expectation for a target near the end of the list. The target words appeared with equal frequency in Positions 3-7 in the lists. Two words always followed each target word on a list.

The set of 50 target words consisted of two groups of two-syllable words. In one group the first syllable of the words was meaningful by itself (such as *can* in *candy*), and in the other group the first syllable was not meaningful (such as *cac* in *cactus*). These two groups each consisted of roughly equal numbers of two syllable words beginning with the phonemes /b/ (9 in *each* group), /s/ (9 in *each* group), and /k/ (7 in *each* group). Each of these initial-phoneme divisions was, in turn, composed of three different first-syllable structures, which were included to provide variety in the sample of initial-syllables used. These structures were

$$\mathrm{C} + \mathrm{V} + \left\{ \begin{matrix} \text{nasal} \\ \text{glide} \\ \text{sibilant} \end{matrix} \right\},$$

C + V + stop consonant, and C+V. Each of these syllable-structure subdivisions contained three words (with the exception, due to experimental length limitations, of the initial phoneme /k/). Thus, the

variables of initial phoneme and initial syllable structure were only roughly factorial.

There were three classes of targets that are or can be considered to be words: two categories of two-syllable words, those having meaningful first syllables and those having nonmeaningful first syllables; and the meaningful first syllables themselves. All three categories had very similar word-frequency distributions as indicated by the Thorndike-Lorge word count. The median frequency score for meaningful first syllables was 19, while the median scores for two-syllable words having meaningful and nonmeaningful first syllables were, respectively, 23 and 17. The means for these same three groups were 38.6, 34.5, 32.3. The modal score for all three groups was 100+ per million (the AA rating in the Thorndike-Lorge count). Thus, it is unlikely that any observed differences in RTs could be attributed to frequency of the target.

All lists were recorded, in a random order, on audio tape. In those lists containing targets, a signal, which was inaudible to the subjects, was placed coincident with the beginning of the target word by manually moving the tape across the playback heads until the onset of the target was located. (Since the same tape was used for all conditions, any signal placement error applied equally and thus was inconsequential.) The signal initiated a timing mechanism in a PDP8/I computer, which stored all latencies. Timing was terminated by the subject's response.

Procedure. The subjects were instructed that they would hear lists of two-syllable words, each list preceded by the specification of a target. They were told that the target would be either a two-syllable word or the first syllable of such a word, and they were informed that the target did not actually occur on all of the lists. The subjects were instructed to press a button (on which they rested an index finger at all times) as soon as they heard the specified target. The stimuli were presented to the subjects over headphones at a rate of approximately .5 sec per word. There was a pause of approximately 3 see between trials.

Design. The subjects were divided into two groups of 20. Each group heard exactly the same tape; the only difference between them was the target for which they listened. Group 1 listened for single-syllable targets and word targets on the lists where, respectively, Group 2 listened for word targets and single-syllable targets. The subjects in each group responded to 25 syllable and 25 word targets. The single-syllable targets for Group 1 were always meaningful words in themselves, for Group 2 they never were.

Results

The mean latencies for the various conditions were obtained for each subject. Table I shows the mean RTs for the experimental conditions. An analysis of variance indicated that the mean RT in the single-syllable target condition (374 msec) was significantly longer than the mean RT in the word condition (341 msec), $F(1, 38) = 29.88$, $p < .001$. The two groups did not significantly differ from each other, $F < 1$. Likewise, the Groups x Target interaction was not significant, $F(1, 38) = 1.19$, $p > .25$. The failure of the interaction suggests that the difference between the single-syllable RTs for meaningful and nonmeaningful initial syllables was not significant. A planned between-subjects t test was performed on the data from these two conditions, confirming this expectation, $t(38) = .95$, $p > .10$.

The RT data were further broken down as

TABLE 1

MEAN REACTION TIMES (MILLISECONDS)

EXPERIMENT I

Group	Target							
	Syllable				Word			
	/s/	/b/	/k/	Mean	/s/	/b/	/k/	Mean
1[a]	395	371	356	381	380	332	329	342
2[b]	387	362	353	366	353	319	350	341

[a] Nonmeaningful first syllable.

[b] Meaningful first syllable.

TABLE 2
MEAN NUMBER OF OMITTED RESPONSES EXPERIMENT I.

Group	Target	
	Syllable	Word
1[a]	2.40	0.36
2[b]	1.20	0.20

[a] Nonmeaningful first syllable.
[b] Meaningful first syllable.

a function of the initial phoneme of the target, /s/, /b/, and /k/. In all cases the syllable target was responded to more slowly than the word target (see Table 1). In addition, the syllable target led to longer RTs for each of the initial-syllable structures that were included in the materials.

Occasionally subjects would fail to respond to the presence of a target. Table 2 presents the mean number of omitted responses in each of the conditions. As can be seen, many more responses were omitted in the syllable case when the initial syllable was nonmeaningful as opposed to when it was meaningful, $t(38)$ 7.3, $p < .001$. The status of the initial syllable had no effect on the number of omitted responses in the word-monitoring condition.

Discussion

The major result of this experiment seems to be quite straightforward and again, from one point of view, somewhat surprising. The RTs to monitor for two-syllable words were shorter than those for the initial syllables of those words. This held even though the information for syllable identity would seem, logically, to be available sooner than the information required for the identity of two-syllable words.

It should be noted that, while statistically reliable, the difference between syllable- and word-monitoring RTs was not enormous, about 30 msec. The mean differences that Savin and Bever observed in their comparisons of phoneme- and syllable-monitoring varied from 43 to 119 msec.

The status of the initial-syllable meaningfulness variable is somewhat uncertain. Meaningful first syllables led to significantly fewer omitted responses than did nonmeaningful first syllables. The RT data showed that the former were responded to more quickly than the latter, but this effect did not reach significance. Generally, RT data are more sensitive than error data, so this pattern of results is somewhat anomalous.

It was considered desirable to replicate the major findings of Savin and Bever and of Experiment 1, and to compare all three types of targets within subjects. Hence, a . second experiment was conducted.

EXPERIMENT II

Method

Subjects. The subjects were 45 undergraduate psychology students serving in partial fulfillment of a course requirement at the University of Texas at Austin.

Materials and procedure. With the exception of the meaningfulness variable (see Design below) and a few other very minor changes, the materials for this experiment were identical to those described in Experiment I. The execution of this experiment was very similar to that of the first experiment with the exceptions of a change in experimenter and a change in the instructions. The latter explained to the subjects that the target specified before each trial could be the initial phoneme of one of the words, as well as an initial syllable or the entire two-syllable word. The timing of the stimuli was the same as that used earlier. All subjects listened to the same tape.

Design. The subjects were divided into three groups of 15. Each group listened for each target (initial phoneme, first syllable, two-syllable word) on approximately one-third of the trials. The target that was specified on any particular list was a phoneme for one group, a syllable for another group, and a word for a third group. The three groups, then, constituted counterbalanced conditions wherein RTs were obtained for the initial phoneme, first syllable, and both syllables of each target word. The initial phonemes, /s/, /b/, and /k/, were also counterbalanced across the groups and targets. The first-syllable meaningfulness variable was included in the present study but it was not perfectly counterbalanced across the other conditions.

Results

The mean RTs were obtained for each subject in each condition under study and

these data were subjected to a three-way analysis of variance: Groups x Target (phoneme, syllable, word) x Initial phoneme of target (/s/, /b/, /k/). Overall, the mean RTs to respond to phoneme, syllable, and word targets were, respectively, 442, 359, and 336 cosec. These means were significantly different, $F(2, 84) = 123.4$, $p < .001$. A planned within-subjects *t* test was performed contrasting the syllable- and word-monitoring conditions. The difference between them was significant, $t(44) = 3.96$, $p < .005$.

The effect due to initial phoneme was significant, $F(2, 84) = 30.9$, $p < .01$. In addition, the initial phoneme effect interacted with the effect due to target type, $F(4, 168) = 2.44$, $p = .05$. Although these two effects interacted, the relative ordering of the RTs for the various targets was the same for each initial phoneme condition. That is, phonemes were responded to most slowly and words most quickly for each of the three initial phonemes.

The main effect due to Groups was not significant ($F < 1$), but this ettect did enter into two significant interactions. The Groups x Target interaction was significant, $F(4, 84) = 5.45$, $p < .01$, as was the three-way interaction of Groups x Target x Initial phoneme, $F(8, 168) = 16.5$, $p < .01$. Again, however, these interactions do not change the pattern of the results since they reflect differing degrees of the main effect across the groups rather than reversals of the main effect. Thus, for each group, the shortest RTs occurred when the target was a word and the longest when the target was a phoneme. The data reflected by the three-way interaction present a similar picture. If one expects that phoneme RTs will be longest, syllable RTs next, and word RTs shortest, then there are 18 pairwise comparisons that can be made in the Group x Target x Initial phoneme interaction. Of these, 15 came out in the expected direction and three in the reverse direction.

Although the first-syllable meaningfulness factor was not completely counterbalanced across the other conditions, a post hoc analysis of the effects of this factor was carried out. The mean RT to monitor for meaningful first syllables was 353 cosec while that for nonmeaningful initial syllables was 369 msec. An analysis of variance performed on these data failed to reach an acceptable level of significance, $F(1, 42) = 3.35$, $p < .10$. As opposed to Experiment I, there was no difference in the number of omitted responses when the target was a meaningful as opposed to a nonmeaningful first syllable.

The principal results of this experiment are shown in Table 3.

TABLE 3

MEAN REACTION TIMES (MILLISECONDS) EXPERIMENT II

Initial phoneme	Target			
	Phoneme	Syllable	Word	Mean
/s/	481	389	336	412
/b/	431	358	316	368
/k/	415	328	325	356
Mean	442	359	336	379

Discussion

The two studies presented above yielded a consistent pattern of results: In both cases the RTs to respond to syllables were longer than those for two-syllable words. The second study showed that, within subjects, the order of RTs, from longest to shortest, was phoneme, syllable, two-syllable word. Warren (1971) has also demonstrated that RTs for words are shorter than those for components of the words. Many of his materials were presented in sentence contexts.

From the observation that single syllables led to shorter RTs than did phonemes, Savin and Bever concluded that phonemes were not real units of perception, but rather were abstract. By parity of argument, we must now conclude that single syllables are not units of perception, but rather that this honor is reserved for the word. This seems to be an unhappy conclusion since it denies that any of the systematic information inside words is

utilized in perception. The sadness is made rather profound by the results of a further experiment carried out by Bever, Savin, and Hurtig, which is briefly mentioned by Bever (1970). In their work, Bever et al presented lists of three-word sentences like (3) to their listeners

(3) Monks ring chimes, cows give milk,
shoes help feet, plants have seeds,
boys like girls,...

The sentences were presented at the rate of one sentence per second. The subjects were told to monitor for either an entire three-word sentence, such as *boys like girls*, or the initial word of a sentence, such as the word *boys*. Bever summarized the results as follows, "Our results so far show that subjects respond consistently faster when they know the entire sentence target than the initial word target..." (p. 25). No further details were given. Again, by parity of argument, we must now reject the word as a unit of speech perception and accept the phrase or clause as basic. But this, of course, is a reductio ad absurdem since there are simply. too many phrases or clauses possible in the language for units at this level to be *the* basic perceptual entities. It is obvious, then, that the entire series of experiments needs reinterpretation.

GENERAL DISCUSSION

One potential tack to take in reinterpreting these studies is to suggest that the results reflect the listener's degree of uncertainty concerning the acoustic identity of the target. Since the phoneme (perceptually) is acoustically different in different syllabic contexts, when a subject is told to listen for a phoneme he is in a state of uncertainty as to the acoustic nature of the target for which he is monitoring. Likewise, there will also be some (lesser) uncertainty in the case of a syllable target. We might predict that, in general, the more uncertainty, the longer the RT since the subject will not know the acoustic structure of the target.

The above "acoustic pattern match" theory appears to account for the data in the experiments under discussion, but it does not. The phoneme used by Savin and Bever in their second experiment was /s/. Likewise, one-third of the targets in the present two experiments began with /s/. It has been shown in tape-splicing experiments with real speech (Harris, 1958) that the acoustic cues for /s/ are invariant across different following vowels. Likewise Heinz and Stevens (1961) have shown with synthetic speech, that "formant transitions do not have an appreciable effect on /s/ ... " (p. 596). In the present studies, the ordering of RTs in the cases where /s/ was the initial phoneme was the same as the overall ordering (i.e., phoneme, syllable, two-syllable word). Thus, one cannot validly account for the ordering in terms of the uncertainty of the acoustic signal.

There is also another, weaker, reason for rejecting as an interpretation of the present results an hypothesis that focuses on the acoustic signal. There was some evidence in the two experiments that the meaning of the first syllable of the two-syllable words had an effect on syllable-monitoring RTs. The subjects were somewhat faster in monitoring for meaningful first syllables than for non-meaningful first syllables. The effect of the meaning variable on RTs was admittedly quite small and thus provides only a tenuous argument against an acoustic interpretation. However,.the two arguments, taken together, do seem to provide convincing reasons for rejecting an acoustic uncertainty hypothesis.

Note that we are not claiming that acoustic factors had nothing to do with any of the results. Both experiments showed that RTs were shorter for /b/ than for /s/ targets, even though the former is more highly encoded. It may be the case that the syllables that began with /s/ were spoken somewhat more slowly and that this acoustic difference can account for the RT difference. There is also the fact, observed here as well as in the Savin and Bever experiments, that the RTs were very short

relative to the stimulus duration. The syllables in Experment I were spoken at a rate of approximately 500 msec/syllable. The overall syllable RTs were 374 msec. Hence the subjects were obviously not waiting to hear the entire syllable before initiating their response. These times can, no doubt, be manipulated by varying the acoustic similarity of the nontarget syllables to the target syllable. Thus, acoustic factors are important in considering some aspects of the experiments.

In accounting for the present results, it does not seem feasible to divide the units of perception into those that are "abstract" and those that are "real". All levels of analysis, from phonetic to semantic, result in entities that are "abstract" in the sense that they are not simply given in the acoustic signal. Some processing of the input must occur for any of these entities to result.

Following Studdert-Kennedy (in press), we can think of speech perception as a transformation of the input signal into a message in terms of stages of analysis: (1) auditory, (2) phonetic, (3) phonological, (4) lexical, syntactic, semantic. All stages but the first result in abstract entities. The processes that result in these entities are not independent; higher level decisions influence those at lower levels. The entities at Stages 2 and 3 can be thought of as phonetic and phonological matrices with both segments and junctures represented. One set of junctures will signal the syllable. The phonetic syllable is a very important perceptual unit resulting from Stage 2 analysis. Studdert-Kennedy has pointed out that this unit serves a number of important functions; for example, it permits the listener to contrast the segments of sound in order to facilitate auditory discrimination. But while this unit serves some important functions, these do not necessarily make the syllable the primary unit at either Stage 2 or Stage 3 processing.

The identification of syllable boundaries is itself determined by constraints at both higher and lower levels. At the higher end, for example, consider the sequence *anicebox*. This sequence can be divided into either *a nice box* or *an ice box*. The structure of the syllables (i.e., which one /n/ goes wills) is determined by which words are accessed in the lexicon. Sentence contexts can be devised so that either segmentation is the most plausible and is the one perceived. The lower level constraints are those on the form of permissible phoneme sequences in English syllables. For example, no syllable can begin with the sequence /kt/. If such a sequence is heard, the listener need not check his lexicon to see whether there is a word containing a syllable beginning with /kt/ since the rules of syllable structure forbid it. If the rules of syllable structure are brought to bear upon the input (clearly an empirical question), then that input must be segmented into subsyllabic units. This follows because the rules are stated in terms of such units.

In sum, we know that phonemes are not, by themselves, the perceptual primes. They arise, in part, through an analysis of higher units. We have argued that it is likewise a mistake to propose that syllables, by themselves, are the perceptual primes. (It should be noted that Savin and Bever did not argue explicitly that syllables had this status. Such a conclusion does seem implicit in much of their discussion, however.) It seems clear that progress on the topic of speech perception will be made only by a laborious working out of the various rules, and their interactions, that determine the entities that arise at the various stages of perception.

At this point, we are still left with the problem of interpreting our results as well as those obtained by Savin and Bever (1970) and by Warren (1971). One rather obvious point by now is that the monitoring task does not tap into the comprehension processes at a level that corresponds to immediate perception. In some earlier papers (e.g., Foss & Lynch, 1969) this "obvious" point was missed. We have argued that the perception of a two-syllable word does not precede the perception of the word's first syllable. However, it is, reasonable to suppose that the larger units may become available to

the listener in the natural course of comprehension while the smaller constituent units do not. By "available" we mean, roughly and imprecisely, that the unit becomes conscious and that an externally observable response can be made contingent upon it. When an item arises in consciousness we will say that it has been identified. Thus, there is a distinction between the perception and the identification of a linguistic unit. Independently, a similar distinction has been mentioned by Warren (1971).

The distinction between perception and identification rests on a particular view of the speech processing mechanisms. We assume that stages like those proposed by Studdert-Kennedy (in press) actually exist that there are-processes that convert the neuroacoustic input into a neurophonetic representation, that there are other processes that transform the entities at the latter level into a representation at the neurophonemic level, and so on. As stated above, these processes do not operate straight through; they are subject to modification by feedback from higher-level analyses. When speech processing has proceeded to the point where the phonetic-to-phonemic transfer function can apply, then speech has been "perceived" at the phonetic level. At the time syllable-structure constraints have applied and lexical look-up takes place, speech has been "perceived" at the level of phonemes (even though not all of the phonemes have been perceived and even though some changes in perception can still occur because of higher level constraints). If the listener's task is to monitor for a particular phoneme, he may not be able to initiate his response contingent upon the outcome of processing at this relatively low level. Processing there is inaccessible to consciousness. This is not an unusual claim. One is likewise unable to get direct access to many other perceptual processes; for example, we are not directly aware of the working of our visual edge detectors but of the figures that are constructed from them.

Generally, listeners do not identify the phonemes or syllables in the speech they perceive. It is uncertain to our intuitions whether we usually identify the words we hear when comprehending connected discourse. We think that we do not always to do so, although task demands can clearly influence this. Phrases and clauses (more accurately, they meanings thereof) seem to be the natural units of identification or awareness. It seems reasonable to, suggest that the RT data observed in the present studies reflect the order in which units are, or can be, identified rather than the order in which they are perceived. Identification of words occurs sooner than identification of syllables; identification of syllables occurs faster than that of phonemes. (The reason for this ordering will be discussed below.)

The assumption that larger units are identified before smaller ones leads to the further idea that the smaller units are identified by fractionating the larger ones. That is, there may be processes for decomposing identified units into their constituents. (A similar suggestion was made by Savin and Bever.) These processes may, to a large extent, be learned. That is, it may not be the case that for purposes of identification one is naturally able to fractionate words into syllables, syllables into phonemes, or perhaps even phrases into words. Thus, young children's institutions about the constituents of words do not seem to be very well developed.

(Parenthetically, it is worth noting that in the case of phoneme-monitoring, some subjects' decisions may occur via an examination of the spelling of the identified lexical item or its first syllable. The letter will then be the unit that is identified and enters into conscious awareness. The introspective reports of some of the subjects in phoneme-monitoring tasks agree with this analysis.)

We will conclude this paper with some rather vague speculations about why the larger units of analysis are the ones that seem to be identified earliest. Our hypothesis is this: When we process input stimuli, one sufficient condition for a unit or construction to enter into aware-

ness is that it leads to a change or an attempted change in long-term memory. Long-term memory can change in a number of different ways. Its contents per se can change, as when one adds an item to some list structure (e.g., the capital of Albania is Tirana). Long-term memory can also change by reorganizing its contents, or by changing the relations among its contents. And, importantly, long-term memory changes when there is a change in some process (e.g., the auditory-to-phonetic transfer function). Note that such changes are, by hypothesis, a sufficient condition for conscious awareness. Other events, many having nothing to do with processing external stimuli, may also lead to consciousness.[2]

Typically, the phonetic and phonological analyses of the input do not themselves lead to awareness, that is, to. identification of the relevant entities extracted by those analyses. The reason why the phonetic and phonological analyses do not lead to awareness is that no change in the listener's long-term memory is required contingent upon processing the phonetic input per se. For example, the listener does not have to reorganize the rules that convert the phonetic representation to the phonemic representation, nor does he have to increase or change his store of phonetic or phonological units. Since this is true, and since no overt action is required as a result of processing at these levels, the phonetic. and phonological units are perceived but not identified.

There is an exception to this typical case: When a listener hears a speaker with an unusual accent, or with a very different fundamental frequency in his speech, he often does become aware of some difficulty or uncertainly in his phonological analyses, through not of the analyses themselves. This is manifested behaviorally by requests to repeat or to slow down, or by not comprehending the message. By hypothesis, what is happening in these cases is that the acoustic-to-phonetic or phonetic-to-phonological transfer functions require some change. That is, a change is being made in the hearer's long-term memory and this leads to a state of awareness. The listener is not aware of the changes themselves, but of a surface manifestation of them, that is, that he is doing additional nonautomatic processing. He is also sometimes aware of the uncertainty in the resulting units: "Did you say *rapid* or *rabid*?" When the rules have been changed, the fact that the accent is being processed then drops out of awareness; its processing becomes automatic.

Some processing eventuates in the meaning of phrases or clauses being perceived. At this level, the perceived entities (and their relations) are almost always identified as well. That is, they arise in conscious awareness. According to our hypothesis, this suggests (but does not entail) that some change in long-term memory has followed upon the perception of those entities. For the most part, this suggestion seems quite plausible. The present analysis appears to break down, however, when one remembers that much of what we hear is either not new, circular, tautologous, or vapid. Should perceiving such utterances lead us to change the contents of our long-term memories and hence lead to identification and conscious awareness? Under our hypothesis, the answer is "yes", since some change in the contents of long-term memory is, in fact, contingent upon concluding that the input is not new, vapid, and so on. These changes occur i n the system of knowledge and beliefs that the listener has about the speaker, the decisions that got him into his current situation, and so on. Roughly speaking, he must update his memory representation of the speaker and of his own past

[2] The present hypothesis bears a close resemblance to the theory that novel stimuli lead to attention and awareness. There is a difference between these views, however. Thus, not all novel stimuli (either types or tokens) lead to awareness, especially if one is processing other inputs (e.g., the dichotic listening situation). In addition, "novelty" is a derived concept. A stimulus is novel *with respect to* the memory system of the observer. Thus a novel stimulus that is identified as well as perceived may be one that leads to an attempted reorganization of memory.

wisdom. Thus, the meanings of the first few sentences of a discourse in which the listener is engaged will always be identified since they are bound to lead to some change in long-term memory. If the change is of the latter sort, it is soon completed, the listener "turns-off," and other events enter into awareness. The memory-change hypothesis has a number of interesting implications; however, they are beyond the scope of this discussion.

The problem of consciousness is, of course, an enormously complex one and it can be discussed from many points of view (for some recent statements by psychologists, see Bindra, 1970; Festinger, Burnham, Ono, & Bamber, 1967: Posner & Boies, 1971; Posner & Warren, 1972: Sperry, 1969, 1970). The present approach concerns the specification of a sufficient condition for conscious experience. Compiling a taxonomy of such sufficient conditions could conceivably be an interesting and tractable task.

In summary, then, the present paper has presented evidence that RTs to detect two-syllable target words are shorter than those to detect the initial syllables of those words and that the latter RTs are shorter than those to detect the initial. phoneme of the syllable. Arguments were presented that these results were not simply due to acoustic factors in the signal. In order to account for the results, a distinction between the perception and identification of linguistic units was introduced. The results were said to reflect the order of identification of the relevant entities rather than the order of their perception. Some speculations were forwarded concerning the determinants of the order of identification of entities. These speculations hinged on the idea that conscious identification occurs whenever a change in long-term memory occurs.

REFERENCES

BEVER, T. G. The influence of speech performance on linguistic structure. In G. B. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North Holland Publishing Co., 1970.

BINDRA, D. The problem of subjective experience: Puzzlement on reading R. W. Sperry's "A modified concept of consciousness." *Psychological Review*, 1970, 77, 581-584.

CHOMSKY, N., & MILLER, G. A. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 11. NY.. Wiley, 1963.

FESTINGER, L., BURNHAM, C. A., ONO, H., & BAMBER, D. Efference and the conscious experience of perception. *Journal of Experimental Psychology Monograph*, 1967, 74, No. 4 (Part 2).

Foss, D. J. On the time course of sentence comprehension. In J. Mehler & F. Bresson (Eds.), *Proceedings of the 1971 C.N.R.S. Psycholinguistics Conference,* impress.

FOSS, D. J., & LYNCH, R. H., JR. Decision processes during sentence comprehension: Effects of surface structure on decision time. *Perception & Psychophysics*, 1969, 5, 145-148.

HARRIS, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, 1, 1-7.

HEINZ, J. M., & STEVENS, K. N. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 1961, 33, 589-596.

JAKOBSON, R.,.& HALLE, M. Phonology in relation to phonetics. In B. Malmberg (Ed.), *Manual of phonetics*. Amsterdam: North Holland Publishing Co., 1968.

KOZEVNIKOV, V. A., & CISTOVIC, L. A. *Rech' Artikuliatsia i vospriiatie*. Moscow-Leningrad : Nauka. Translated as *Speech: Articulation and perception:* Washington: Clearinghouse for Federal Scientific and Technical Information, Joint Publications Research Service, 1965, 30, 543.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, 74, 431-461.

POSNER, M., & BOIES, S. J. Components of attention. *Psychological Review*, 1971, 78, 391-408.

POSNER, M., & WARREN, R. E. Traces, concepts, and conscious constructions. In A. W. Melton & E. *Martin (Eds.), Coding processes inhuman memory*. Washington: Winston, 1972.

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

SPERRY, R. W. A modified concept of consciousness. *Psychological Review*, 1969, 76, 532-536.

SPERRY, R. W. An objective approach to subjective experience: Further explanation of a hypothesis. *Psychological Review*, 1970, 77, 585-590.

STUDDERT-KENNEDY, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics*, Vol. 12. The Hague: Mouton, in press.

WARREN, R. M. Identification times for phonemic components of graded complexity and forspelling of speech. *Perception & Psychophysics*, 1971, 9, 345-349.

(Received October 6, 1972)