PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Role of Syllables in Speech Processing: Infant and Adult Data [and Discussion]

J. Mehler and R. W. Hayes

Phil. Trans. R. Soc. Lond. B 1981 **295**, 333-352
doi: 10.1098/rstb.1981.0144

Printed in Great Britain

The role of syllables in speech processing: infant and adult data

By J. Mehler

Laboratoire de Psychologie, École des Hautes Études en Sciences Sociales (*C.N.R.S*),
54 *Boulevard Raspail*, 75006 *Paris*, *France.*

An empirical account is offered of some of the constants that infants and adults appear to use in processing speech-like stimuli. From investigations carried out in recent years, it seems that syllable-like sequences act as minimal accessing devices in speech processing. *S*s are aware in real time of syllabic structure in words and respond differently to words with the same initial three phonemes if the segmental one is CV/... and the other CVC/....

Likewise, infants seem to be aware that a 'good' syllable must have at least one alternation if it is composed of more than one phoneme. When the segment is only one phoneme long, its status is necessarily somewhere between that of the phoneme and the syllable.

An important problem that arises with the syllable is that it is an unlikely device for speech acquisition. Indeed, there are a few thousand syllables and the attribution of a given token to a type is far from obvious. Even if physical invariants for syllables in contexts were to be found, the task facing the child still remains one of sorting thousands of types from many more tokens. Issues concerning acquisition versus stable performance will be addressed to further constrain possible models. In addition, I try to show that even though information processing models are useful tools for describing synchronic sections of organisms, the elements that can account for development will have to be uncovered in neighbouring branches.

Traditionally, psycholinguistic research has invested the bulk of its efforts into uncovering the units used in speech processing. Although it is currently fashionable to claim that such work is pointless since it has no very clear outcome, many of the more meaningful advances in the field have come from projects whose framework included the problem of processing units (see, for example, Liberman *et al.* 1974; Stevens & Blumstein 1978). Nevertheless, it must be acknowledged, in the light of the work of the Gestalt psychologists, that there are also dangers in oversimplification. Thus, it would be false to use the basic parameters of physical optics, for example, to account for visual perception. In relatively recent work, Gibson and colleagues have tried to demonstrate that accounts of the perceptual processes can best be understood in the context of the ecological whole. But this 'direct registration approach', important as it may be, neglects the computational processes used in perceiving stimuli as varied as sentences. As Ullman (1980) states, 'When an exhaustive enumeration becomes prohibitive, processes and rules of formation would offer an advantage over the direct coupling of input–output pairs' (p. 375). Indeed, two different groups of detractors from the study of units exist. The first maintain that perceptive processes are global and cannot be analysed; but their claim cannot be accepted for speech since it would be tantamount to abandoning any hope of a solution. The second assert that computational processes must be abandoned in favour of mapping functions. However, as mentioned earlier, such a position is untenable for cases where exhaustive enumeration is not viable. Thus, I conclude that there is no alternative to the analytic study of speech processing as characterized by what has already been undertaken in the area.

My own approach to this problem is anchored in a computational conception of cognitive processes, even though the empirical concern of the work that I have carried out with my colleagues falls within the range of what Marr & Poggio (1977) call the algorithmic level of information-handling systems. Furthermore, we cover both speech in the adult stable state as well as dispositions for speech in the newborn. In fact, it is quite possible that information processing models may benefit from being developed with speech acquisition in mind. As Chiat (1979) claims, '...the child's primary task is to break up the speech chain, to isolate meaningful units, rather than to determine how such units are distinguished from one another' (p. 592). This belief that segmenting the speech chain into units is as important as classifying their contents may well be correct and should contribute towards the study of adult speech processing. We can pursue this work at two different levels.

On the first level, linguists search for units that in spite of being minimal can account for the phonological components of grammar. Such units express the phonological structure of languages as economically as possible. In this context, the phoneme, the distinctive feature, etc., are the most interesting candidates. The syllable has also been introduced to increase the descriptive accuracy of the phonological component of grammars. Indeed, Selkirk (1978), Halle & Vergnaud (1976), Kahn (1976) and others look on the syllable as a useful tool in the study of phonology.

On the second level, concern with speech perception and speech comprehension requires an autonomous search for the most fruitful units used by *S*s during speech perception and comprehension. I shall concentrate on perceptual aspects exclusively. Although this has already been done by Miller & Nicely (1955), the general assumption is that the tools that are good in one explanatory domain, such as phonology, should suffice in another, speech perception. Although such a position may turn out to be correct and has been defended at other levels (see Bresnan 1978), we must accept the fact, pragmatically, that there is a vast difference between the insight displayed by formalists and that of experimentalists. Furthermore, rather than induce the latter to adopt the tools of the former, this fact should encourage each of them to seek as much autonomy as possible since it is not at all obvious, *a priori*, where the junction, if junction there be, will occur.

We know that phonemes and distinctive features play an important, if not a principal, role in speech perception. What we do not know is whether the language user, when he or she listens to speech, is activated sequentially or in parallel by the phonemes themselves or by their distinctive features. There is considerable scepticism among certain psycholinguists concerning such a purely bottom-up version of speech perception. This scepticism may be due, perhaps, in part, to the obvious fact that there is no understanding of how such a transducing device may operate. But we cannot refute a hypothesis because circumstances prevent us from putting it into effect. We also know that context is influential in determining what is heard. Furthermore, a number of observations have led some theorists to imagine that top-down processes cannot be neglected in the construction of a theory of speech perception. Without denying their existence, I shall attempt to show that speech perception is largely data driven. In addition, I believe that our understanding of how a top-down mechanism operates hinges upon the determination of the most basic units in speech processing.

Savin & Bever (1970) proposed that the syllable might be the basic unit in speech processing on the basis of an experiment in which the syllable was most accessible to the response-triggering system. Indeed, they demonstrated that *S*s press a button faster to a syllable target than for its initial phoneme. From this result they concluded that phonemes are not perceived directly but are derived from analysis of the syllabic unit, which is perceptually the primary device. Needless to say, the logic behind this conclusion came under attack from many different quarters. Foss & Swinney (1973) showed that *S*s respond to words faster than to syllables. By induction, if clauses elicit faster responses than words it might be possible to argue that the clause is the primary perceptual unit. Indeed, such data were reported by McNeill & Lindig (1973). To extract arguments from such counterintuitive grounds, Foss & Swinney brought in a distinction between perception and identification. They claimed that '...smaller units are identified by fractioning larger units' (p. 254). As I shall argue later, this view may be partly correct although it requires qualification. The distinction introduced by Foss & Swinney is a proper one provided that empirical facts can be uncovered about perceptual processes over and above those relating to identification. McNeill & Lindig claimed that all results obtained with the Savin & Bever technique could be accounted for in terms of the linguistic levels of the target item and the items in the search list. If the levels were the same, the response times (t_r) had to be low; as the linguistic distance increased, the t_r increased also. The authors thus concluded that the method could not reveal anything about perceptual units in speech perception.

Healy & Cutting (1976) compared phoneme and syllable monitoring and concluded that it is the relative ease with which a target is identified that determines whether a given syllable or given phoneme will be monitored fastest. Their conclusion was that '...it seems best not to consider either the phoneme *or* the syllable as the basic perceptual unit but rather to consider the phoneme *and* the syllable as linguistic entities equally basic to speech perception' (p. 82). However, as Mehler *et al.* (1981 *b*) have argued, there are several problems with McNeill & Lindig's experiments as well as with those of Healy & Cutting. First, '...as the authors themselves admit, *S*s when given targets specified as syllables, are unable to disregard the accompanying vowel. In fact, Woods and Day (1975) have shown this experimentally. Thus, the conditions that the authors define as phoneme monitoring are in reality syllable monitoring. McNeill and Lindig apparently used the same vowel environment for all items in the phoneme list condition' (pp. 419–430). This was shown to have an important effect both by Swinney & Prather (1980) and by Mills (1980). If a syllable or its initial phoneme are monitored with varying degrees of vowel predictability, it can be shown that syllable monitoring is no faster than phoneme monitoring when the vowel of the latter is matched to the one used in the target or when the list has only one vowel in all items. Alternatively, as in Healy & Cutting's experiment, the phonemes used, namely those that can be uttered in isolation, have an ambiguous status. The phoneme /a/ for instance, can be a vowel, a bound morpheme or even a word. In all likelihood, *S*s construct a syllabic representation of the target and the closer their representation is to that of the critical item, the faster they respond. Indeed, as Mills claims, '...the closer the match is between the listener's expectancies and the stimulus the *S*s actually hear, the faster they are able to identify the stimulus.... Therefore faster RTs [t_r] should be obtained for syllables than for phonemes, because for syllables the contexts of the initial phonemes of the target and stimulus always match, whereas for phonemes, the contexts of the target and stimulus rarely, if ever, match' (p. 76).

As I shall demonstrate, there is a sense in which it is unreasonable to take exception to Mills's mismatch hypothesis. Nevertheless, I hope that it will soon be obvious that this view has certain limitations in that it dismisses the more or less privileged status of the syllable over other speech segments.

The syllable

In presenting the syllable as a plausible candidate for on-line processing of the speech signal, several general arguments, none entirely conclusive, will be put forward. Following that, I shall explore several alternatives and in particular that of the direct mapping of acoustic signals onto lexical items.

Most of the phonemes in natural languages can be both syllables and phonemes. In the case of a vowel like /*a*/, such a speech sound can function as a phoneme or as a syllable as in the word *a*//*gitated*. It can also act as a bound morpheme or a lexical item in both English and French. There are, however, some phonemes that are only phonemes, i.e. stop consonants. It is interesting to note that it is in the nature of such stops that they cannot be produced either naturally or by splicing outside of a vocalic context. In a recent experiment, Blumstein & Stevens (1980) report (p. 660),

> ...listeners are able to extract both consonant and vowel information from these brief stimuli that can be as short as one glottal pulse (together with an initial burst) [a glottal pulse has a duration of roughly 10 ms]. One way of interpreting this finding is that the brief stimulus signals the identity of the *syllable*, which is processed by the listener as a unitary percept or a single event. Having identified the syllable, the listener is then able to indicate its consonantal and vocalic components.

In commenting upon the systematic fashion in which acoustic phoneticians have tried to determine invariants for consonants, Stevens & Blumstein (1978) state (p. 661),

> ...for the most part, this work failed to find a set of acoustic properties that are invariant for a particular place of articulation, independent of the following vowel... One way of looking at the auditory processing of the stimulus is to imagine that a number of auditory detectors respond selectively to different properties, such as compactness of onset spectrum, abruptness of onset, etc. This constellation of properties leads to the identification of the syllable.

Other studies have also lent credibility, in a more or less incidental fashion, to the syllable as a kind of Gestalt-like unit of the acoustic speech signal. Howell (1978), for instance, and Howell & Darwin (1977), in experiments where *S*s hear stimuli that they must discriminate, have found that the detail of the acoustic signal is not preserved for much over 400 ms. Howell concludes that the auditory level appears to operate upon syllable-sized time windows and that matching '...depended on the properties of the syllable rather than on the properties of the phoneme' (p. 294). Massaro (1972, 1975), using a masking technique, found that the auditory image was available for a duration roughly comparable with that of the mean syllable length. Huggins (1964), with a switching technique, found that the greatest interference was obtained when the duration corresponded roughly to that of a syllable. Hall & Blumstein (1978) showed that adaptation and test stimuli must share a number of properties defined in syllable-sized segments when adapting putative detectors. Indeed, if adaptation is to occur, they must have not only the same phonetic and syllable structure but also the same number of syllables.

Other incidental observations could be cited in this context. For instance, many of the well organized recognition routines rely on the syllable (Mermelstein 1975; Fujimura 1975; Smith 1977; Vaissière1980). An obvious exception is Klatt's (1980) SCRIBE proposal, to which I shall return at greater length below. For the time being, we must consider the overall problem of speech recognition in both the child and the adult. The listener confronted with Skinner's very famous sentence 'Anna Mary candy lights since imp pulp lay things', can hear it as 'An American delights in simple play things'. Miller (1963) gives another example, 'The good can decay many ways', or 'The good candy came anyway', and Chomsky & Miller (1963) cite one of the many examples that exist in French: 'Gal, amant de la Reine alla (tour magnanime)', or 'Gallamment de l'arène à la Tour Magne à Nimes'. Which version of these ambiguous utterances does the listener actually hear? It all depends on the way that sentences are segmented. It is in this sense that segmentation processes in fluent speech are, necessarily, the means by which words can be isolated. Of course, in an exhaustive search like that proposed by Klatt, all possible versions are presumably considered before one is retained. However, although this model may work in an abstract sense, it is probably incapable of accounting for either the facts of on-line processing, lexical access or even speech acquisition since it operates exclusively on words as primary acoustic patterns. One thing does, however, appear obvious: that is that infants, and even young children, hear speech largely in terms of sentences rather than isolated words.

We shall now consider some of the more salient problems attached to a model that contains no segmenting routine. Certainly, as far as the pre-perceptual acoustic store is concerned, it cannot be particularly efficient, since it will be unable to retain the initial segments of a long word if lexical access has not taken place during the first few milliseconds after onset. But, unless very efficient, access can never take place under these circumstances. In addition, if a non-word were presented, *S*s would achieve no representation whatsoever. Finally, as mentioned earlier, the number of lexical items in language is enormous and when exhaustive enumeration becomes too great, a table look-up approach is a poor solution, particularly when time constraints apply. Thus it seems more plausible that speech may be segmented and the results of such segmentation used to attain an intermediate representation, half-way between the acoustic signal and the lexical item. Unfortunately, we remain at a loss to describe the intermediate level. What we propose is that the syllable is the output of a segmenting device that scans the acoustic speech signal and that syllables are used, in conjunction with contextual information, to access the lexicon. Furthermore, we claim that syllables are useful devices for infants during language acquisition.

As mentioned earlier, speech segmentation models often have one thing in common and that is that they largely aim at isolating syllable-sized units by searching for the minima or the maxima in the amplitude envelope. Given that the universal property of languages is the alternation of consonantal and vocalic information, this segmentation procedure is quite satisfactory. However, if we postulate a special processing status for the syllable, then we must be able to come up with supporting empirical facts. I shall now present some data on how words presented in lists are accessed.

Words in lists

Before presenting data gathered in our and other laboratories, I must mention that we work with French stimuli. However, to some extent, we are convinced that this does not radically change the pattern of results and that they should be easily reproduced in English. We have at least one concrete example that can be cited in passing. Mehler *et al.* (1978) studied phoneme monitoring of initial stop consonants of a word in a French sentence in a setting comparable with that used by Newman & Dell (1978) with English sentences. Figure 1 gives t_r's for monitoring phonemes in a word preceded by a one-, two- or three-syllable word. We plotted the results in both experiments (for comparable targets) and, as can be seen, they are almost identical, within a few milliseconds of each other. I shall return to these results in the next section, but would merely like to point out that they have confirmed our belief that processing of the kind of information we are studying, at the level under consideration, is probably largely independent of any specific language. However, I shall not digress any further and shall return to the processing of words in lists.

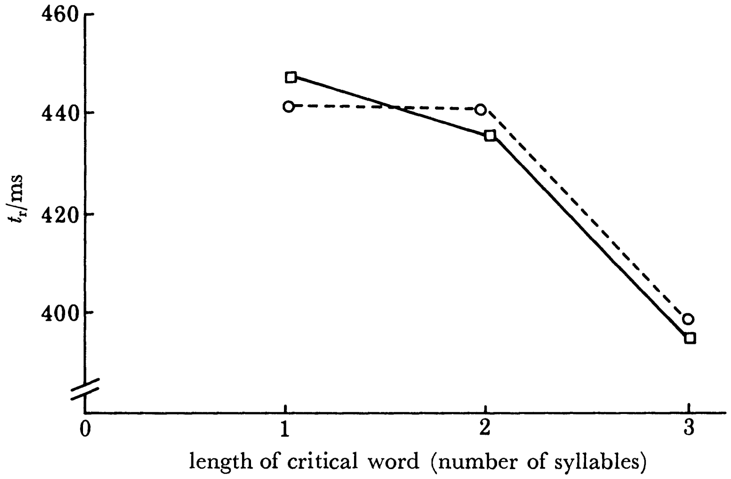

FIGURE 1. Comparison of the comparable data by Newman & Dell (1978) (- - -) and Mehler *et al.* (1978) (—) on French and English sentences.

Major theories, for example those of Morton (1969), Forster (1976), Marslen-Wilson & Tyler (1980), are relatively neutral as to the actual acoustic or phonetic correlates used in accessing the lexicon. Foss & Blank (1980) reversing the position initially espoused by Foss & Swinney, believe that there are two ways for *S*s to respond to part of a word in a phoneme monitoring task. In one, *S*s gain access to the phonological code through a direct table 'look-up' procedure. In this case they use the information in the lexicon to make a response. In the other, *S*s use the phonetic structure of the word to generate a response without having accessed the lexicon. As I shall show shortly, if *S*s are called on to respond to a phoneme in a target word they use either the phonetic or the phonological code.

My claim is that *Ss* can respond to a phoneme in a target through syllabic segmentation or through the phonological code. This hypothesis thus differs from that of Foss & Blank in that we claim that syllables, or syllable-sized segments, are used by *S*s for all purposes, i.e. making a monitoring response, accessing the lexicon, etc. It is also possible that *S*s can guess the word that they are going to hear in very predictable sentences, before they have received any part of the corresponding acoustic signal. In such cases, they may use the phonological code to respond to the target – the phonological code being represented in the lexical entry for the target.

If the syllable is generated by the segmenting device it still remains a mystery how this device operates and how it classifies its output. Consider the words *carotte* and *cartable*. Although both words share the first three phonemes, the initial syllable of *carotte* is /ca/ while the initial syllable of *cartable* is /car/. Does this mean that the segmenting device isolates different segments for both these words? If the length of the segments isolated is approximately identical, will their classification be the same? These and other questions are raised in the course of an experiment carried out by Mehler *et al.* (1981*a*), in which it is shown that *S*s respond differently to pairs of words sharing the first three phonemes but having different syllable structure. If the target is made up of the first two phonemes only, say /ca/, when we use the example cited above, t_r's are fast for *carotte* and slow for *cartable*. However, if the target is /car/ the t_r's are in inverse order. Figure 2 clearly shows the interaction between type of target and type of word as found in this experiment. Targets were presented visually and *S*s were instructed to respond to a word in a list where the initial segment matched the target. These results indicate that the speed at which *S*s respond to a target is a function of that unit's syllabic status in the word. The prototypical response time is about 360 ms. If, as is often proposed, the preparation time for a response is on the order of 100 ms or more *S*s trigger a response on the basis of a stimulus whose duration is roughly comparable to that of a syllable. As we shall see below, it thus appears likely that *S*s in our experiment responded before they had accessed the lexicon. But how can we explain these results? Notice that although the target is the same, and the representation that the *S*s have in mind is stable, the response times to a /CV/... or /CVC/... type word vary. McNeill & Lindig's linguistic level-matching hypothesis thus cannot account for these results and nor can Healy & Cutting's view based on matching and identifiability of target items. Things are slightly different for Mills's match-mismatch hypothesis.

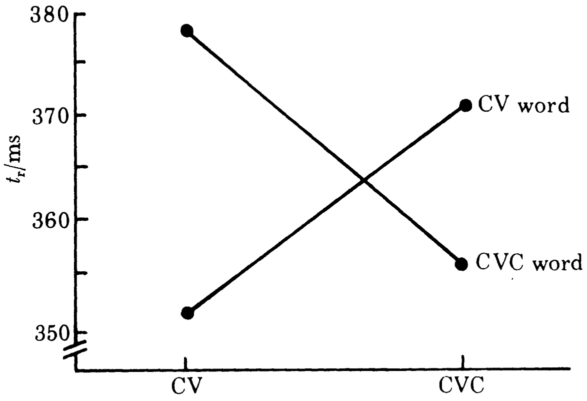

FIGURE 2. Mean reaction time for CV and CVC words as a function of target type.

Let us suppose that after having read the target, /ca/, *S*s store it in purely abstract terms. Our results would then suggest that syllables are processed in real time and that the classificatory status of a /CV/... item is closer to the abstract representation of /CV/... target than that of a /CVC/... item. If the target receives an acoustic representation rather than an abstract one, the same argument could be made (although the nature of the matching would undoubtedly seem more comprehensible). Thus, the only conclusion that might be drawn from our experiment is that syllables may be units in on-line speech processing. Indeed although our findings are compatible with those of Mills they are also orthogonal to them in that they only suggest that the syllable or an acoustic correlate is the output of the segmentation device. This result is understandable in the light of claims by linguists such as Abercrombie (1967) or Fudge (1969), who claim that the duration of a vowel depends on the number of consonants that follow it in the same syllable. Syllable initial consonants are shorter than syllable final ones, although some systematic distributional properties at the feature level also distinguish syllable initial and syllable final consonants (obviously, this latter claim cannot apply to our sequences).

To gain a better understanding of the representation of syllables, we carried out another experiment in which we tested a hypothesis derived from the idea that *S*s segment speech signals syllabically. If subjects respond to a /VC/... target, they should respond to it faster when the response item is contained within a single syllable than when the /V/ is in one syllable and the /C/ in another, i.e. responds to /al/ in *palace* or *palmier*. Using a procedure quite similar to that presented earlier, we found the results shown in table 1.

Table 1

target type	CV word	CVC word
V	615	614
VC	704	658

Careful inspection of these results suggests that *S*s respond to a target that corresponds to the first /V/ in a target word beginning with a consonant at a speed roughly identical for the different target words used, although they take twice as long in this situation compared with their speed when they are called on to respond to the first phoneme in the words. Why is this so? Perhaps because if what elicits the response is the syllable or one of its initial segments, this unit is immediately available as part of the output of the segmenting device (an intermediary syllabi-phonetic code). On the other hand, if it is only a part of such a unit that is used to trigger the response, the output of the intermediary code has to be analysed into components and classified to match with the target.

If we take the response to the /VC/... type target, a reliable and larger difference (54 ms) than could have been predicted appears. Indeed, when *S*s respond to a /CVC/ type word, all the information they need is contained in the first output of the segmenting device, whereas if they respond to a /CV/ /C.../ type word they need two such outputs before they have the information needed to trigger a response. Perhaps these results can be explained best if we assume that *S*s respond after having accessed the lexical item. Of course, this would entail that there is also a syllabic organization of the lexical code and consequently that when items are within a single syllable, t_r's will be faster than when they are distributed between two syllables. Long t_r's make this hypothesis likely, since in over 600 ms *S*s have heard most of the word.

Morton & Lucio (unpublished) have studied the repetition of syllables just before having to report a word. *S*s heard 200 bisyllabic words among which were 30 critical triplets. The relevant property of the triplets was that the first syllable of the first words and the second syllable of the second word were the two syllables of the third word in that order. For example, if the first word was *conquer* and the second *progress*, then the critical item was *congress*. The test word was presented in a noise context such that there was a 50% intelligibility score. Context words according to group were also heard with noise present for one group and not for the other. The results indicate that both groups perform better on the test word when it is preceded by a facilitating context at the syllabic level. As Morton & Lucio themselves state, one way to account for this result is to claim that there are units at the level of the syllable that are prior to the structures responsible for the recognition of words. Furthermore, their result is compatible with the interpretation offered for the experiment of Mehler *et al.*

So far, the data appear to indicate that the syllable is used as part of the bottom-up analysis of speech, at least as far as words in lists are concerned. However, other data contradict this view. Rubin *et al.* (1976) showed that initial phonemes elicit faster response times in words than in non-words by using a procedure that made subjects search in parallel for initial /b/ and /s/ targets by pressing one of two keys for each type of target. The /s/ type targets were not scored. This procedure is not standard since most of the other investigators working in the area have used one target and one response key. An interesting feature of this experiment resides in the fact that the *S*s heard only *monosyllabic* items. Thus, to some extent, phenomena pertaining to lexical access may explain their results. Indeed if we compare the two blocks of stimuli used by Rubin *et al.* in their experiment:

Table 2

	word	non-word
phoneme	347	346
syllable	285	281

KEEJ, NUG, LAN, NAEN, SIM, DAJ;
JUT, LEG, SIN, RUG, WELL, RUN;

it is obvious that the relative frequency of syllables corresponding to words and non-words is very different. In view of the amount of existing data demonstrating that access is facilitated by the frequency of the item, it is clear that the use of monosyllables makes aspects of bottom-up analysis of the signal as well as aspects of lexical access inextricable. Thus, Segui *et al.* (1981) have looked into the role of the lexical status of the target items when *S*s monitor for initial syllables or initial phonemes in those items. All items are *bisyllabic* target items (either words or non-words), and all targets started with stop consonants. Our results contrast sharply with those of Rubin *et al.* We find, as shown in table 2, that under the phoneme monitoring condition, words elicit a response at 347 ms, and non-words after 346 ms. There is no difference in the t_r's depending on lexical status. For Rubin *et al.* the equivalent t_r's for words were 593 ms and non-words 644 ms. This discrepancy calls for two observations. First, the fact that we found no difference of lexical status may have been due to our using bisyllabic items, which allowed us to distinguish the data-driven aspects of signal analysis from lexical access. Given their procedure, Rubin *et al.* may not have been able to do this. Secondly, the very complex method that they used may partly account for the difference in t_r between our *S*s and theirs. Our *S*s were on an average twice as fast as those of Rubin *et al.* Furthermore, our *S*s, under the syllabic monitoring condition, respond to words in 285 ms and non-words in 281 ms. This exaggerates the difference between the two experiments, but again no trace of any effect of lexical status of the item on t_r is uncovered.

Thus, if we consider the results reported this far and the results recently reported by Mills (1980), the case for the syllable as a bottom-up unit in signal analysis is quite strong. Mills shows that a one-syllable utterance when emitted in isolation is responded to faster when that syllable is given as a target than when the same syllable is included in a two- or three-syllable word. Thus, when *S*s are given /can/ as a target they respond to the word *can* faster than to /can/ in the words *candlelight* or *candle*. Likewise, *S*s respond faster to the word /can/ than to the same word produced by splicing the words *can/dle* or *can/dlelight*. To cite Mills (p. 534):

> ...The results of this experiment showed the perceptual consequences of coarticulation information that crosses the stimulus boundary. As predicted on the basis of the target-stimulus mis-match hypothesis when the target was one syllable, subjects were able to recognize the stimuli that were spoken as isolated syllables faster than those that were spoken as part of two syllable utterances or three syllable utterances... Thus, these results show that the ease of identifying a given stimulus is not only determined by the absolute acoustic characteristics of the stimuli itself but also by its similarity to the other syllables that make up the string.

Mills, furthermore, observed that the amount of coarticulation in the first syllable of a three-syllable word is no different from that in a two-syllable word. Thus, these results, rather than being a hindrance for a data-driven system of the analysis of the signal, show that coarticulation yields information as to the syllabic length of the word from which the first syllable was taken. That is, *S*s have information as to the morphemic status of the item that they are responding to. Latency to respond to a syllabic target /can/ is roughly as fast when spliced from *candle* or *candlelight* but slower than when *can* is in itself a word. This observation will be an important one in our presentation of data on syllable use during lexical access in sentence perception.

Finally, and to summarize our position, we uphold the hypothesis that the syllable is probably the output of the segmenting device operating upon the acoustic signal. The syllable is then used to access the lexicon. Monosyllabic words access the lexicon automatically, while polysyllabic items use the first syllable to make tentative access or even two or more syllables according to context probably along the lines of some revised version of Marslen-Wilson & Tyler's cohort system. For the time being, it is not clear whether or not there is an alternative to this hypothesis – whether or not the initial segment of an acoustic signal, say the first 20 ms suffices to identify a potential syllable as in the work reported by Blumstein & Stevens (1980); and, on the basis of this identification, whether the beginning of the next syllable can again be identified with some precision. However, since all the data collected so far concern stimuli that are either words or non-words presented in lists, it is not clear how much of this account can be used in the context of perception of sentences, which is what ultimately has to be accounted for.

The syllable in sentence perception

As mentioned earlier, the phoneme-monitoring technique has been one of the principal tools employed in on-line studies of sentence perception. Over time a number of important findings have emerged. After the pioneering work done by Foss (1969) and Foss & Lynch (1969), several observations have been made suggesting that the data may reveal processes related to lexical access. For instance, Morton & Long (1976) showed that if the context makes a word very predictable, the initial phoneme of that word elicits a faster response than if the context makes the word unpredictable. At that time, Morton & Long interpreted their finding much as Foss & Swinney had done before them by saying that the phoneme is responded to after lexical access has occurred. According to a logogen model, '...the amount of work involved is a function of context (that is of the material being processed) and it is not affected by the word actually being recognized by the separate word identification system' (p. 49). The prototypical test sentences were like (1 *a*, *b*).

(*a*) *branch*
(1) A sparrow sat on the whistling a few shrill notes to welcome the dawn.
(*b*) *bed*

If we consider Morton & Long's data, we see that for words starting with a plosive in high or low transitional probability contexts, the first are monitored with a 386 ms latency, while the second elicit a response after 460 ms. These results, added to several others like those of Foss & Swinney, lend some support to the view that phonemes are responded to after lexical access has occurred. Indeed, in Morton's model context lowers the threshold for the *branch* logogen and therefore would have an effect in (1*a*) but would not affect the *bed* logogen at all. Thus there would be no facilitation due to the context in (1*b*). This facilitation can only be at the lexical level, and consequently only after having accessed the word can the *S*s access the phonological code and be aware that the word begins with a *b*. Thus, the longer the accessing, the slower the monitoring response. It therefore comes as no surprise that the monitoring response is given with greater speed to *branch* than to *bed*.

Table 3

	on	after
word	475	525
non-word	481	626

In Mehler *et al.* (1978) as well as in Newman & Dell (1978), the syllabic length of the word preceding the target-bearing one seems to be an excellent indicator of phoneme-monitoring times. This can be taken to mean that in on-line sentence processing it is important to access words for at least two reasons: first, because, as has often been suggested, if accessing occurs on-line it allows the rather confident detection of word boundaries; secondly, the words so accessed can be combined into higher computational structures according to the 'obligatory operation' processing proposed by Marslen-Wilson & Tyler. The results reported suggest that for short words *S*s may be partly overloaded, or still engaged in processing that makes the next syllable, namely the target-bearing one, less available for response. For longer words, the redundancy makes it plausible that *S*s having already accessed the item in the lexicon respond to the target because they are free from other commitments and have no difficulty in detecting the end of the word.

More recently, Foss & Blank (1980) have presented a series of results in an experiment in which *S*s had to respond to sentences like (2*a*, *b*).

(*a*) *government*
(2) At the end of last year the prepared a lengthy report on birth control.
(*b*) *gatabond*

*S*s had either to respond to the /g/ in *government* or *gatabond* or the /p/ in *prepared*. At first sight, the results are quite intriguing. In table 3 we see the t_r's to a phoneme in a target word or non-word or to a phoneme following one of these. As can be seen, there is hardly any difference in the monitoring times for responses when the target phoneme is in the word and responses where the target is in the non-word. *Government*, moreover, should be relatively easy to access, whereas *gatabond* is not. Nevertheless, there is no phoneme-monitoring difference at all.

However, and in contrast, when the target is on *prepared*, reaction times are much faster when following a word rather than a non-word. These results have to be considered in conjunction with the finding that phoneme-monitoring times are not affected by the frequency of occurrence of the target-carrying word in sentences like (3*a*, *b*).

(*a*) *t*eacher
(3) Yesterday afternoon the — *b*orrowed the article from the reference library.
(*b*) *t*utor

Indeed, no overall frequency effect was found on the critical word. It should be noted, however, that in all likelihood (3*a*) has a high transition probability towards *teacher* while sentence (3*b*) has a lower transition probability towards (*tutor*). A fairly marked effect was however, detected on the target following the critical pair, i.e. the response is faster after a high-probability word than after a low-probability one. This result might be due to concomitant task demands. Whereas Morton & Long asked their *S*s for recall of all the test sentences, Foss *et al.* asked for a comprehension test only. Thus, borrowing 20 of the stimulus sentences used by Morton & Long, they tried to replicate these findings by using the task demands as a parameter. One group had to recall the stimulus sentence by rote while the other group had to recognize the test sentence in a recognition test. Their results can be seen in table 4. In reporting their results, Foss and Blank find that they corroborate Morton & Long's findings under their task conditions. To cite these authors, '...experiments that manipulate transitional probability yield results consistent with the hypothesis that *S*s respond to the target after retrieving the target-bearing word. In contrast, experiments that manipulate inherent characteristics of the target-bearing word (e.g. lexical status or frequency) yield results indicating that *S*s respond prior to retrieving the word' (p. 17). But what is the reason for this? What is the difference between high and low transition probability and relative frequency in affecting a phoneme monitoring response? Foss & Blank account for this in terms of the Dual Code Hypothesis. What this hypothesis suggests is that *S*s employ either a phonetic or a phonological code in speech perception. In the first they compute a set or bundle of phonetic features. In the second, a phonological code (probably what is stored in memory to be able to utter that word) becomes available. The use of either code can yield a response under the right conditions. Thus, if the context is such that it strongly suggests a word, *S*s respond in terms of the phonological rather than the phonetic code. Although some version of this hypothesis may very well turn out to be correct, a convincing way of accounting for the available data seems none the less to require the syllable. Indeed, we reanalysed Morton & Long's data and showed that they mostly use monosyllabic pairs. When polysyllabic pairs are used (unfortunately very few), the effect almost disappears. Mixed cases are very interesting. Indeed, when a pair contains a polysyllabic and a monosyllabic word, a greater difference in t_r (115 ms) ensues when the polysyllabic word is in the high transitional probability context than when it is the monosyllabic item that is in

Table 4

	context	
	probable	improbable
recognition	405	449
recall	409	456

the same frame (64 ms). We might speculate that when *S*s are given the polysyllable in high transitional probability they respond very rapidly in terms of the first syllable, but are slowed down on the monosyllabic item for reasons that have to do with lexical access being necessary but slow because of the low transition probability. In contrast, if lexical access is accelerated, as for instance when the monosyllabic word is the one to appear in the high transition probability context, t_r should be shorter because the monosyllable, in this case, is accessed faster. Even if context does not affect the polysyllabic items, we can interpret the differences. However, if the context also affects the availability of the first syllable of the polysyllabic item, the direction of this effect could only be to increase the predicted differences. If this interpretation is correct, we can account for the effect found by Morton & Long as well as that found by Foss & Blank when they replicated the Morton & Long experiment by borrowing 20 of its sentences, most of them monosyllabic. As was initially pointed out by Cutler & Norris (1979) this possibility is quite real since in their other experiments Foss & Blank used polysyllabic words. Furthermore, the difference may explain why *S*s do not seem to respond faster to a high-frequency target-bearing item compared with a low-frequency target-bearing item. Indeed, it must be that polysyllabic items receive responses from the syllabi-phonetic code that may be less affected by the frequency of the overall item. This interpretation must be maintained with great caution since Norris (1981) has recently carried out an interesting experiment in which this hypothesis is overtly tested. In experiment 9 of his dissertation, Norris tests monosyllabic against polysyllabic targets in high and low transitional probability contexts. He finds a significant main effect of transitional probability, although, as might have been predicted, this effect is considerably greater for short words than for long. However, the interaction between length and transition probability is not significant. Following this, Norris carried out another experiment in which the initial phoneme of the word was preserved while the rest was changed to construct a non-word of the same syllabic length. Likewise, a second pair of non-words that did not share the initial phoneme was created. In this experiment, an effect contingent upon the transition probability of the critical word was again found for the word materials. No effect was found for the non-words irrespective of whether they shared the initial phoneme with the predictable word or not. As Norris comments, '...the fact that the latencies for non-word targets are almost identical to latencies of low transitional probability words provides considerable support for the race model proposed by Cutler and Norris. Responses to non-word targets can only be based on the phonological analysis of the stimulus and therefore the similarity between non-word latencies and latencies to low transition probability targets would imply that responses to low transition probability targets are also based on a phonological analysis alone as Cutler and Norris predict' (unpublished).

Although Norris's experiments are very impressive, they do not seem to close the case. First, a number of his long words were bi-morphemic as in *pass*port or *post*card and it is possible that such words lead to the accessing of *pass* and *post* respectively. Furthermore, a number of his bisyllabic items may be shorter at least psycholinguistically, i.e. compare *turtle* with *turnip*: obviously the former seems less polysyllabic than the latter. It is very possible that these two factors, plus the fact that some contexts are much less constraining that other, have generated data like those found by Norris.

In our laboratory, Segui, Frauenfelder and Mehler are currently exploring this issue somewhat further in an experiment in which the targets are always polysyllabic. The high transition probability item is then compared within the same frame with a low transition probability item. A third item, that shares the first syllable of the high-probability target, is also introduced into the frame. This experiment will clarify whether syllables, as opposed to phonemes, are detached from their lexical whole during scanning of the acoustic signal.

The syllable as a unit in speech processing must be looked on as a speculative device even if for the time being it seems to account for many of the data rather neatly. Many issues still remain open. For instance, supposing that syllabogens are devices just prior to logogens, are they facilitated by context or are only logogens, as such, subject to top-down constraints? Other issues concern the determination of the actual syllabogenic invariants as they are represented in the environment. Indeed, one may speculate that the output of the segmenting device might be close to the acoustic signal or that its output is abstract and indifferent, say, to allophonic variations that are not critical for the language in question. These and other questions will have to be settled before the syllable as a unit of on line processing can be taken seriously. Be that as it may, a segment of syllabic size does seem to be actively engaged in all the processing data gathered in the on-line studies reported here. The phoneme, which undoubtedly plays an important classificatory role and a basic articulatory role, has not been shown to be used in on-line processing of sentences unless it is used at a level so basic as to be entirely opaque to behavioural measurement. Thus, all in all, if Morton's logogen model is adapted to take syllables into consideration, then a more or less open picture of sentence processing emerges.

Many arguments have been advanced against syllables in on-line processing. For instance, it is claimed that if the segmenting device goes blindly from minima to minima in the signal (given a time constant) it will segment chunks that do not correspond directly to the morpheme structure of the sentence. Furthermore, it may cross the boundary between words in a chunk. However, if the results of Mehler *et al.* hold as well as those of Mills, a dynamic model of sentence perception based on syllables, syllables used for accessing the lexicon, lexical items providing clues to word endings and beginnings of the next syllable, as well as contentive items for higher level integration, then the above-mentioned objections may be disposed of. Indeed, our finding that it is the actual syllable rather than the phonemic sequence that is the chunk most available for response may be important in that it makes access to the lexicon more efficient. Even in a 'cohort' type model, the acoustic signal is segmented and initial syllables are used to access the lexicon much as Marslen-Wilson & Tyler (1980) suggest for a different segment. Essentially, word recognition occurs after sufficient acoustic–phonetic information has become available to allow the distinction between that item and all other items that provide the same initial acoustic–phonetic information. What we propose is that such decisions are taken syllable by syllable rather than in a continuous time sample, or phoneme by phoneme. Once words have been accessed, further processing is carried out to compute the sentence form and content. The finding by Mills suggests that *S*s may find it far easier to determine syllables that are lexical entries, from syllables that are part of a long word. If we look at both of these observations, it is very probable that the kind of problems raised above do not apply in the actual course of sentence processing, given that more cues are available to guide this performance than was previously acknowledged.

Whether the syllable is the tool we are proposing for sentence processing or not, it is interesting to evaluate its role in speech acquisition since there are a great many critical observations that could make or break the syllable's plausibility. Obviously, the most interesting (because it is the most economical and elegant) theory in this context is that of a 'table look-up' for features or for phonemes that might be mediated by specialized detectors. Eimas *et al.* (1971) were the first to entertain this hypothesis. Ten years later, we must acknowledge that, on balance, the data are not overwhelmingly favourable to that position. Furthermore, it is very difficult to understand the means by which such detectors could be calibrated to take into consideration the language-specific values of parameters such as VOT (i.e. in English, Thai or French). Again the argument that such a simple 'table look-up' view could solve many problems with only a few detectors, is not viable for syllables. Indeed, languages use many thousands of syllables and this very fact makes the syllable difficult to conceive as an element in language acquisition. But the problem is certainly no different from that of lexical acquisition or, for that matter, facial recognition or a great many other perceptual cognitive abilities with which we are so much at ease that we minimize them once their acquisition has taken place.

If we look at a completely different set of observations, we see that Liberman *et al.* (1974) reported that young children can and do monitor syllables when they are unable to tap to phonemes. In addition, Morais *et al.* (1979) have shown that illiterates behave similarly, i.e. they have no difficulty in monitoring for syllables but are unable to monitor phonemes. These results suggest that the syllable may be the first segment that is used in acquisition and that the availability of phonemes might turn out to be a result of specific training.

I am fully aware of the speculative nature of the hypothesis that the syllable is the most basic segment available for speech processing but its neatness in accounting for all the available data, as well as its predictivé value in explaining the breakdown of Morton & Long's data, must be regarded as promising.

But what is the nature and genesis of the syllabogen device? A first and obvious remark is that it must vary widely from language to language. Newman (1952) and Menzerath (1950) have both analysed a number of languages for distribution of vowels and consonants. They found that Italian is made up largely of open CV or VC syllables, while German is constituted of CVC-type closed syllables. French and English are intermediate in their syllable form. Thus, learning a language could mean that certain basic forms, forms that vary from language to language, have to be mastered.

Conclusion in the light of some developmental data

For over a decade we have known that human infants can discriminate syllables that differ only in their initial stop consonant (see Eimas *et al.* 1971; Trehub 1976). Furthermore, infants seem capable of distinguishing syllables whose medial or final phonemes differ (Jusczyk 1977; Jusczyk & Thompson1978). Many other studies could be cited in this context but the critical issues remain unanswered, i.e. whether these studies indicate the existence of specialized feature detectors and whether infants are sensitive to syllables as a primary source of speech signal analysis

In so far as the specialized feature detectors are concerned, the data remain unclear. Nonetheless, as Jusczyk (1981) has stated,

> ...the results from studies of the discrimination of speech sounds do not provide definitive evidence about the existence of a specialized speech mode of perception for infants. There are indications that infants engage in similar processing for speech and certain nonspeech signals. At the same time, recent investigations demonstrating the existence of context effects

> upon the infant's discrimination of speech suggest the existence of processing abilities designed to cope with the complex interactions which can occur between speech cues. Yet, whether these abilities are of phonetic or psychophysical origin is still unknown.

Regardless of what the answer may be, and even if specialized detectors are uncovered, speech processing by infants may turn out to rely on procedures similar to those used by adults. Little is known about their segmenting faculties, but recent experiments by Demany (1981) have shown that infants have streaming constraints similar to those of adults. Bertoncini & Mehler (1981) have been able to demonstrate that infants display greater discrimination for well-formed syllables than for chains of phonemes.

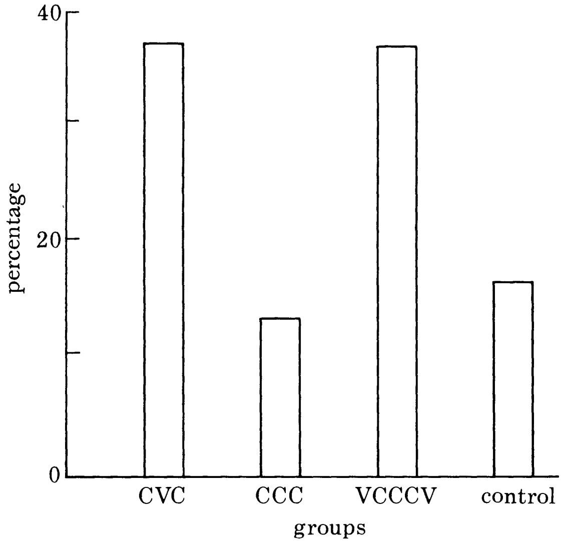

FIGURE 3. Dishabituation performance by four groups of very young infants on CVC, CCC and VCCCV stimuli.

In an experiment aimed at testing the use of syllables in speech processing, Mehler & Bertoncini make the hypothesis that infants segment speech on the basis of syllables. This hypothesis is difficult to test directly but it is possible to test infants' sensitivity to possible and impossible syllables. As stated by Jakobson & Waugh (1979), '...it is the initial sequential contiguity of consonants and vowels which plays the main role in their interrelation within any given language' (p. 86). Jakobson & Waugh mention some apparent counterexamples from Korlak and Bella Cool, i.e. 'vtvt' or 'ktkt'. But Bell (1970) claims that releases of transitional voicoids are always present even in examples such as those and that the problem is at the level of phonetic transcription. Thus, we can make the hypothesis that for natural languages sequences like $C_1C_2C_3$ cannot be words. Thus if infants are sensitive to natural syllables they should consider a syllable like C_1VC_2 to be different from a syllable like C_2VC_1 but should be neutral to a similar physical change when it occurs in non-syllables such as $C_1C_xC_2$ and $C_2C_xC_1$. As shown in figure 3, our results are in line with our expectations. Details of these experiments can be found in Bertoncini & Mehler(1981) and in Mehler *et al.* (1981*b*).

Granted that infants have a segmenting device that emits syllables, the status of such segments in speech acquisition remains unclear. The great attraction of phonemes over syllables as units in speech processing is that they can, at least hypothetically, be imagined as processed by specialized detectors. For syllables their number is so great and their structure so different from one language to the next, that such a hypothesis is somewhat compromised. Therefore, when we propose syllables (and, as has been argued by others, phonemes also) as devices used in speech processing, an acquisition mechanism must be included in the proposal. This is not necessarily a disadvantage. As F. H. C. Crick (1979) once stated (pp. 181, 188),

> ...there are some human abilities that appear to me to defeat our present understanding. We sense there is something difficult to explain, but it seems almost impossible to state clearly what the difficulty is. This suggests that our entire way of thinking about such problems may be incorrect. In the forefront of the problems I would put perception... It seems certain that we need to consider theories dealing directly with the processing of information.

I intend to raise some of these issues here.

In studies of perception, two polar oppositions appear to coexist. On the one hand, we have systems that are sensitive to environmental invariance in energy distribution. For instance, suppose that von Bekesy's (1960) theory of pitch perception were correct, it could then be stated that the maximum of the travelling wave that emerges in the basilar membrane as a function of environmental stimulation works on a particular detector that is the transducer of a given frequency (or pitch, as the case may be). Likewise, similar accounts can be put forward for retinal transduction of retinal parameters, etc. Granted, all the problems would not be solved in the framework of such accounts but a sizeable part of those pertaining to perception might be clarified, at any rate for areas where explanations of this kind prove correct.

But much of perception has not yielded to this kind of account. In fact, perceptual constancies appear to be more pervasive than exceptional. Boring (1942) describes the phenomenon of constancy, with *S*s perceiving the stimulus not so much according to its physical properties but rather as they think it should be. As Gibson (1979) has so well expressed the matter for size constancy, '...The size of the object only becomes less *definite* with distance, not smaller' (p. 160). He goes on to state that '...the implication of this result, I now believe, is that certain invariant ratios were picked up unawares by the observers and that the size of the retinal image went unnoticed. No matter how far the object was, it intercepted or occluded the same number of textured elements of the ground. This is an important ratio...' (p. 160).

Although several *ad hoc* accounts exist for the constancies observed for different parameters, little is really understood. In addition, Bower (1977) for example, has claimed that infants extract constancies without having any, or almost any, intercourse with the objects perceived. Thus, how can they have any knowledge of their properties when the properties are necessary to gain the knowledge in the first place? Many views have been advanced in the area of visual perception, but little is known about the constancies in hearing and speech perception. It is true that the typical observation, that large variations in a parameter leave phoneme perception unchanged, might be taken as a case of perceptual constancy, but a great many opinions exist concerning the perceptual reality of the phenomenon itself (as opposed to a more profound classification of signals). Be that as it may, the issue of constancies has rarely been raised in psycholinguistics.

One exception is given by Kuhl (1979), who tested infants' discrimination of vowels while several concomitant parameters (such as F_0, pitch contours) varied in a fashion that was not pertinent to the discrimination itself. Kuhl used a training paradigm with infants who were only 6 months old and who had no difficulty in maintaining the relevant acquistion in the face of other changes in the signal. In other studies, Kuhl was also able to demonstrate that infants do not need specific training to attain the vowel constancy discovered in her initial training paradigm.

Other constancy effects have been studied at the level of the phoneme across a number of different vowel contexts (see Jusczyk 1981). But the results, if not negative, are at the best confusing. To quote Jusczyk,

> One implication of this approach is that, as Bertoncini and Mehler (1979) have suggested, the syllable serves as the basic unit of speech perception for the infant. Thus perceptual constancy for phonetic segments may be a later development arising through the child's detection of certain regularities that exist between the various syllabic units. These regularities may lie in the acoustic or even articulatory correlates of the syllable.

We are, of course, in total agreement with Jusczyk but his statement raises more problems than it solves. For instance, if an infant given a syllable opens a template through which such a syllable will be recognized on a future occasion, what are the properties that the child will store in the template? What are the criteria for opening future templates given that the infant has already allocated certain templates to given acoustic events? By what means will the infant, given that one of his templates is being stimulated, also record properties that are ancillary, though not critical, to the syllable?

In conclusion, I shall add two points. First, the syllable as a segment must continue to be explored as a device that may generate addresses in terms of which the lexicon is constructed. Obviously, since infants are born without a lexicon they are obliged to open entries in which they will be able to locate the contents. Furthermore, they must label the entries for future retrieval. It is probable that the syllable is used in just this way. Secondly, if our claim concerning the non-perceptual reality of the phoneme is actually borne out, the question of the origin of the phoneme (and, for that matter, the distinctive feature) will have to be dealt with. A possible answer is that each of these elements has reality in articulation but that, as with motor acts, insight into them is poor. We are no more capable of giving an adequate description of the movements that we engage in during a frequent motor act (walking up a staircase) than about a unit of speech production. However, observation of behaviour in others does yield clues that can be used for an organized description of motor acts in general.

The writing of this paper was made possible by the generous hospitality of L. Michaelis and C. Michaelis at L'Harmas. I am grateful for a grant from the Délégation Générale à la Recherche Scientifique et Technique to assist in the preparation of this article.

References (Mehler)

Abercrombie, D. 1967 *Elements of general phonetics.* Edinburgh University Press.

Bekesy, G. von 1960 *Experiments in learning.* New York: McGraw-Hill.

Bell, A. 1970 *Syllabic consonants.* (Stanford University Working Papers on Language Universals, no. 4.)

Bertoncini, J. & Mehler, J. 1981 Syllables as units in infants' speech perception. *Infant Behav. Dev.* **4**, 1.

Blumstein, S. E. & Stevens, K. N. 1980 Perceptual invariance and the onset spectra for stop consonants in different vowel environments. *J. acoust. Soc. Am.* **67**, 648–662.

Boring, E. G. 1942 *Sensation and perception in the history of experimental psychology.* New York: Appleton Century Crofts.

Bower, T. G. 1977 *A primer of infant development.* San Francisco. W. H. Freeman.

Bresnan, J. 1978 A realistic, informational grammar. In *Linguistic theory and psychological reality* (ed. M. Halle, J. Bresnan & G. Miller). Cambridge, Massachusetts: M.I.T. Press.

Chiat, S. 1979 The role of the word in phonological development. *Linguistics* **17**, 591–610.

Chomsky, N. & Miller, G. A. 1963 Introduction to the formal analysis of natural languages. In *Handbook of mathetical psychology*, vol. 2 (ed. D. Luce, G. Bush & E. Galanter). New York: Wiley.

Crick, F. H. C. 1979 Thinking about the brain. *Scient. Am.* **241** (3), 181–188.

Cutler, A. & Norris, D. 1979 Monitoring sentence comprehension. In *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (ed. W. E. Cooper & E. C. T. Walker). Hillsdale, New Jersey: L.E.A.

Demany, L. 1981 Auditory stream segregation in infancy. *Infant Behav. Dev.* (In the press.)

Eimas, P. D., Siqueland, E. R., Jusczyk, P. & Vigorito, J. 1971 Speech perception in infants, *Science, N.Y.* **171**, 303–306.

Forster, K. 1976 Accessing the mental lexicon. In *Explorations in the biology of language* (ed. E. Walker). Vermont: Bradford Books.

Foss, D. J. 1969 Decision process during sentence comprehension: effects of lexical item difficulty and position upon decision times. *J. verb. Learn. verb. Behav.* **8**, 457–462.

Foss, D. J. & Blank, M. A. 1980 Identifying the speech codes. *Cogn. Psychol.* **12**, 1–31.

Foss, D. J. & Lynch, R. H. 1969 Decision processes during sentence comprehension: effects of surface structure on decision times. *Percept. Psychophys.* **5**, 145–148.

Foss, D. J. & Swinney, D. 1973 On the psychological reality of the phoneme. Perception, identification and consciousness. *J. verb. Learn. verb. Behav.* **12**, 246–257.

Fudge, E. C. 1969 Syllables. *J. Ling.* **5**, 253–286.

Fujimura, O. 1975 Syllable as a unit of speech recognition. *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-23**, 82–87.

Gibson, J. J. 1979 *The ecological approach to perception.* Boston: Houghton Mifflin.

Halle, M. & Vergnaud, J. R. 1976 Formal phonology. Unpublished lecture notes, Department of Linguistics, M.I.T.

Hall, L. & Blumstein, S. E. 1978 The effect of syllabic stress and syllable organization on the identification of speech sounds. *Percept. Psychophys.* **24**, 137–144.

Healy, A. & Cutting, J. E. 1976 Units of speech perception: phoneme and syllable. *J. verb. Learn. verb. Behav.* **15**, 73–83.

Howell, P. 1978 Syllabic and phonemic representations for short term memory and speech stimuli. *Percept. Psychophys.* **24**, 296–500.

Howell, P. & Darwin, J. C. 1977 Some properties of auditory memory for rapid formant transition. *Memory Cogn.* **5**, 700–708.

Huggins, A. W. F. 1964 Distortion of the temporal patterns of speech: interruption and alternation. *J. acoust. Soc. Am.* **36**, 1055–1064.

Jakobson, R. & Waugh, L. R. 1979 *The sound shape of language.* Hassocks, Sussex: Harvester Press.

Jusczyk, P. 1977 Perception of syllable-final stop consonants by two month old infants. *Percept. Psychophys.* **21**, 450–454.

Jusczyk, P. 1981 Auditory versus phonetic coding of speech signals during infancy. In *Proceedings of the C.N.R.S. Conference*, Paris, 15–18 June. (In the press.)

Jusczyk, P. & Thompson, E. 1978 Perception of a phonetic contrast in multisyllabic utterances by two month old infants. *Percept. Psychophys.* **23**, 105–109.

Kahn, D. 1979 *Syllable-based generalization in English.* Bloomington: Indiana University Linguistic Club.

Klatt, D. H. 1980 Speech perception: a model of acoustic phonetic analysis and lexical access. In *Perception and production of fluent speech* (ed. R. A. Cole). Princeton, N.J.: Erlbaum.

Kuhl, P. K. 1979 Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *J. acoust. Soc. Am.* **66**, 1668–1679.

Liberman, I. Y., Shankweiler, D., Fischer, F. W. & Carter, B. 1974 Reading and the awareness of linguistic segments. *J. exp. Child Psychol.* **18**, 201–212.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. 1967 Perception of the speech code. *Psychol. Rev.* **74**, 431–461.

Marr, D. & Poggio, T. 1977 From understanding computation to understanding neural circuitry. *Neurosci. Res. Program Bull.* **15**, 470–488.

Marslen Wilson, W. & Tyler, L. 1980 The temporal structure of spoken language understanding. *Cognition* **8**, 1–71.

Massaro, D. W. 1972 Perceptual images, processing time and preperceptual units in auditory perception. *Psychol. Rev.* **79**, 124–145.

Massaro, D. W. 1975 Backward recognition masking. *J. acoust. Soc. Am.* **58**, 1059–1065.

Mehler, J., Segui, J. & Carey, P. 1978 Tails of words: monitoring ambiguity. *J. verb. Learn. verb. Behav.* **17**, 29–37.

Mehler, J. Dommergues, J. Y. & Frauenfelder, U. 1981*a* The syllable's role in speech segmentation. *J. verb. Learn. verb. Behav.* (In the press.)

Mehler, J., Segui, J. & Frauenfelder, U. 1981*b* The role of syllable in language acquisition and perception. In *The cognitive representation of speech* (ed. T. F. Myers, J. Laver & J. Anderson), (*Advances in Psychology Series*). Amsterdam and New York: North-Holland. (In the press.)

Menzerath, P. 1950 Typology of languages. *J. acoust. Soc. Am.* **22**, 698–701.

Mermelstein, P. 1975 Automatic segmentation of speech into syllabic units. *J. acoust. Soc. Am.* **58**, 880–883.

Miller, G. 1963 Introduction to the formal analysis of natural languages. In *Handbook of mathematical psychology*, vol. 2 (ed. D. Luce, G. Bush & E. Galanter). New York: Wiley.

Miller, G. & Niceley, P. E. 1955 An analysis of perceptual confusions among some english consonants. *J. acoust. Soc. Am.* **27**, 338–352.

Mills, C. A. 1980 Effects of context on reaction time to phonemes. *J. verb. Learn. verb. Behav.* **19**, 75–83.

Morais, J., Cary, L., Alegria, J. & Bertelson, P. 1979 Does awareness of speech as a sequence of phones arise spontaneously? *Cognition* **7**, 323–331.

Morton, J. 1969 Interaction of information in word recognition. *Psychol. Rev.* **76**, 165–178.

Morton, J. & Long, J. 1976 Effects of word transitional probability on phoneme identification. *J. verb. Learn. verb. Behav.* **15**, 43–51.

McNeil, D. & Lindig, K. 1973 The perceptual reality of phonemes, syllables, words and sentences. *J. verb. Learn. verb. Behav.* **12**, 419–430.

Newman, E. B. 1952 The pattern of vowels and consonants in various languages. *Am. J. Psychol.* **3**, 369–379.

Newman, J. E. & Dell, G. S. 1978 The phonological value of phoneme monitoring: a critique of some ambiguity effects. *J. verb. Learn. verb. Behav.* **17**, 35–374.

Norris, D. 1981 Ph.D. thesis.

Rubin, P. E., Turvey, M. & Van Gelder, P. 1976 Initial phonemes are detected faster in spoken words than in spoken non-words. *Percept. Psychophys.* **19**, 394–398.

Savin, H. & Bever, T. 1970 The nonperceptual reality of the phoneme. *J. verb. Learn. verb. Behav.* **9**, 295–302.

Segui, J., Frauenfelder, U. & Mehler, J. 1981 Phoneme monitoring, syllable monitoring and lexical access. *Br. J. Psychol.* (In the press.)

Selkirk, E. O. 1978 The French foot: on the status of the mute 'e'. *J. Fr. Ling.* no. 1.

Smith, A. R. 1977 Word hypothesization for large vocabulary speech understanding systems. Ph.D. thesis, Carnegie Mellon University, Pittsburgh.

Stevens, K. N. & Blumstein, S. E. 1978 Invariant cues for place of articulation in stop consonants. *J. acoust. Soc. Am.* **64**, 1358–1368.

Swinney, D. & Prather, P. 1980 Phonemic identification in a phoneme monitoring experiment: the variable role of uncertainty about vowel contexts. *Percept. Psychophys.* **27**, 104–110.

Trehub, S. E. 1976 The discrimination of foreign speech contrasts by adults and infants. *Child Dev.* **47**, 466–472.

Ullmann, S. 1980 Against direct perception. *Behav. Brain Sci.* **3**, 373–381.

Vaissière, J. 1980 Speech recognition as modes of speech perception. In *The cognitive representation of speech* (ed. T. Myers, J. Laver & J. Anderson), (*Advances in Psychology Series*). Amsterdam and New York: North-Holland.

Woods, C. C. & Day, R. S. 1975 Failure of selective attention to phonetic segments in consonant–vowel syllables. *Percept. Psychophys.* **17**, 346–349.

Discussion

R. W. Hayes (*Department of Psychology, North East London Polytechnic, London, U.K.*). In describing, at the end of his paper, his experiment on neonate infants' sucking responses to various combinations of speech sounds, Professor Mehler presented evidence illustrating the fact that when the infant heard a combination of three consonants, CCC, it habituated more than it did when it heard either the consonant–vowel combination CVC or VCCCV. This finding was interpreted as being due to the lack of alternation in the CCC stimulus, but the habituation difference obtained might have been due to the different physical parameters of the contrasted stimulus groups, rather than being due to 'alternation' and 'non-alternation' *per se*. This follows from the fact that the unvoiced plosive and unvoiced fricative consonants [p], [t] and [s] apparently used in the experiment would have had a considerably higher frequency (hertz), and also a lower intensity than the [u] vowels used in the CVC and VCCCV examples quoted. The different rates of habituation of response found in this study could therefore have been due to a basic difference in frequency and intensity in the content of the all-consonant (CCC) against partly vowel (CVC, VCCCV) groups of stimuli, rather than alternation or non-alternation as such. It may prove useful to consider this point in future work.