INFANT BEHAVIOR AND DEVELOPMENT **4**, 247-260 (1981)

Syllables as Units in Infant Speech Perception*

JOSIANE BERTONCINI
Centre des Processus Cognitifs et du Langage

JACQUES MEHLER
Centre Nationale de la Recherche Scientifique

The aim of the research we have carried out is to assess the role of the syllable in the processing of speech in the very young infant. We used three kinds of stimuli: syllabic, non-syllabic and syllabic-synthetic sequences. These were presented to infants who were less than 2 months old in an habituation-dishabituation paradigm. Results indicate that the syllable-like stimuli are discriminated better than the non-syllable-like stimuli even though the physical change from the habituation to the dishabituation stimuli was always the same. We interpret our results as favoring a view according to which the syllable is the natural unit of speech segmentation and processing.

Si les phonéticiens ont prêté attention au travail des lèvres et non à celui du pharynx, ce n'est pas parce que l'un des facteurs en jeu s'est montré plus important que l'autre. Il faut se rendre compte que lorsque la physiologie des sons refuse d'avoir recours à d'autres disciplines, elle est incapable d'établir la hiérarchie des facteurs en jeu. Par conséquent, si en classifiant les sons du langage les phonéticiens ont pris en considération le facteur labial et non pas le facteur pharyngien, c'est uniquement parce que le premier était plus accessible à l'observation que le second. (R. Jakobson, 1976, p. 28)

*This paper and the research described therein were carried out with the support of a grant from the Harry Frank Guggenheim Foundation to Jacques Mehler. We would also like to express our gratitude to the MIT Research Laboratory in Electronics, to Ken Stevens and in particular to J. Bernstein and B. Delgutte, who were instrumental in preparing all the tapes used in the experiment. Reprint requests should be addressed to Jacques Mehler, Laboratoire de Psychologie, 54, Boulevard Raspail, 75006—Paris, France.

One of the main purposes of developmental speech perception is to discover acoustic invariants for processing units. Such a quest has proven to be a rather difficult one. To a large extent, the study of speech is special in that speakers can also act as listeners and vice versa. This has led some theorists to assume that units of production and reception are corresponding entities (see Halle & Stevens, 1964; Liberman, 1967, for a defense of analysis by synthesis and the motor theory of speech perception respectively). However, whether theorists believe in the articulatory nature of units of speech perception or whether they believe in the acoustic nature of such segments, all of them are engaged in a search for units that will provide the greatest descriptive and explanatory range. But as R. Jakobson (1976) has made clear in the epigram above, the quest for units is often a matter of arbitrary assignment.

Attempts to characterize the invariant segments responsible for the perception of speech sounds regardless of context have run into rather serious problems. Thus, the phoneme segment, on which most of acoustic phonetics has concentrated, remains very difficult to describe in terms of physical invariants. Some of the major difficulties encountered in trying to describe the acoustic invariance of the phoneme have been reviewed by Liberman (1970), Studdert-Kennedy (1979), Mermelstein (1978), etc. Although such problems may be quite normal in work in this relatively new field, it still would appear fair to claim that its investigators have failed in their attempt to describe a viable mechanism for segmenting the speech signal into phonemes. Some researchers, moreover, have gone as far as to question whether such a segmenting mechanism is even necessary in understanding the way in which language users normally process the speech signal. Studdert-Kennedy (1979) maintains:

> . . . . My inclination is to suppose that the preliminary auditory segmentation (if any) is syllabic rather than phonemic and that within-syllable segmentation may often be synonymous with classification. (p. 63)

Though some rather more complex ways of segmenting the speech signal are currently under consideration among acoustic phoneticians, the issues in developmental psycholinguistics are somewhat special in that the only proposal that has been fully developed on the infant's disposition for speech is one based upon the phoneme. In fact, Eimas' (1975) hypothesis claims that infants come equipped with a set of innate feature detectors that insure the contrast between, say, /b/ and /p/ as well as that distinguishing /d/ from /t/ or /g/ from /k/. The general hypothesis according to which infants possess feature detectors for all possible sounds in natural language has two advantages over all other formulations; namely, it captures both the specificity of language and the existence of a putative speech mode. In fact, the advantage of Eimas' hypothesis resides both in the fact that it is capable of giving a natural account of the above-mentioned points and in the fact that it has triggered some of the more interesting investigations into infants' dispositions for speech perception. Unfortunately, the data collected over the last few years does not appear to lend much support to the feature detector hypothesis. Thus, Cutting and Rosner (1974) have shown that infants treat speech and non-speech sounds (having the relevant underlying acoustic similarity) in much the same fashion. Similar findings with adults as well as with infants have been reported by Miller et al. (1976), Pisoni and Lazarus (1974), Mehler and Bertoncini (1979) and others. Along a different line of evidence, the notion that the adaptation studies of Eimas and Corbit (1978) favor a detector interpretation has recently been challenged by Bailey (1975) and by Simon and Studdert-Kennedy (1978). Thus, given the lack of an empirical sub-stratum of feature detectors, we may go even further and ask whether it is likely that adults segment the continuous flow of speech into phonemes or parts of phonemes.

On the basis of the evidence reviewed above, we will formulate a developmental speech hypothesis that, though seemingly promising, has not yet been empirically grounded.

We consider continuous speech as being processed in segments or chunks of a few hundred milliseconds each. Such segmentation helps to discharge some of the memory load that might otherwise ensue. The delimitation of the chunks may be obtained by operating upon the maxima and minima in the wave envelope of the speech flow signal. This segmentation procedure results in chunks which contain consonantal and vocalic information. If subjects do indeed segment in this way, they should arrive at something close to the consonant-vowels (CV's) of speech. Such segments or chunks could conceivably function as inputs to fixed templates, much as the logogen in Morton's lexical access model, where such logogens are associated with acoustic input signals. Thus, CV units are the potential syllables of any natural language. Infants construct templates for the CV's as well as for the V's in their language. Furthermore, they must undoubtedly employ procedures that, of course, need not imply decomposition, which allow them to distinguish whether a sound differs sufficiently from that already coded in an existing template to warrant a new template that matches it. We will not speculate here on the nature of the processing, since we are not currently in possession of enough information. What we can say is that, if infants behave as we have suggested above, they come rather close to dividing the speech signal into units close to the syllables of language. This, of course, raises the possibility that instead of learning a few phonemes before learning their language, children may be faced with the task of learning a great number of syllabic templates. But psycholinguists invariably prefer the more economical model and this may be seen as a critique of the hypothesis being presented. It should be noted, however, that although in semantics and lexicology the initial temptation is also to search for primitives, the success in doing so has been very limited. There is no reason, therefore, for us to expect such primitives to be simpler in phonology. (For a recent overall view of compositional models, see Fodor's (1975) *Language of Thought* and his arguments against definitions contained therein.)

In order to test the speculative hypotheses above, we ran the experiment that will be described in greater detail below and that was to show whether infants, in fact, do process speech in the ways we have indicated. Although it is difficult to imagine a direct test of their segmenting ability, testing the usage and exploitation by infants of the regularities entailed by syllables is easily conceivable.

In all natural language, the alternation between consonant and vowel is an essential part of their underlying structure (see Chomsky & Halle, 1968; Newman, 1952; Menzerath, 1950). All natural languages have words that basically respect an alternation between consonant and vowel. There are very rare examples of three-phoneme words with a consonant-consonant-consonant (CCC) structure. Thus there must be some constraint on alternation that allows a given amount of distortion from consonant-vowel-consonant (CVC), but not too much. Given this regularity, we conceived of testing a child first with either $C_1V_xC_2$ or $C_1C_xC_2$ syllabic or non-syllabic sequences, respectively, and then testing his or her discrimination with stimuli that are constituted by the same sounds in a different order, namely $C_2V_xC_1$ or $C_2C_xC_1$, respectively. In both cases the physical change in the sequence was exactly the same. For adults, however, a change in the order of consonants in a CVC results in an entirely different acoustic image, while an equivalent change in CCC's is far less obvious. In a subsidiary experiment, 10 adult subjects were required to discriminate the same stimuli as those used with the infants. They were supposed to respond "same" or "different" after each pair of stimuli they heard. Although the difference between the response times to CVC and CCC does not reach significance, there is a tendency in all *S*s to respond "different" correctly faster to CVC stimuli than to CCC (367 msec and 394 msec respectively). Almost all the errors occurred in response to CCC stimuli. If acoustic contrasts are more salient in CVC's than in CCC's, then we can predict that infants should discriminate syllabic-like stimuli better than non-syllabic ones. Nonetheless, even if the results we predicted could be obtained, the interpretation would remain slightly unclear because it could be that a CVC is intrinsically, and for acoustic reasons, far more discriminable than an equivalent CCC. Our prediction, however, was that if a CCC sequence was placed in a vocalic context, for instance, $VC_1C_xC_2V$, the stimulus of the sequence should become entirely discriminable. This is because the segmenting device can now split the consonant cluster into a vowel with one, or at most, two consonants attached. Natural languages offer many examples of such consonant clusters that do not produce major difficulties for the speech perception device. Therefore, infants who are unable to discriminate a CCC sequence should, on the other hand, demonstrate some discrimination for the VCCCV sequence.

EXPERIMENT

In a pilot study, we tested infants in an habituation-like paradigm using a non-nutritive sucking technique. Twenty infants under six weeks of age were divided into two equal groups. One group received a CVC type stimulus (rak/kar) while the other a CCC stimulus (šsf/fšs). The rate of presentation of the stimuli was determined by the infants' sucking rate. For both groups, the stimuli were changed by permuting the initial and final consonants: $C_1V_XC_2/C_2V_XC_1$, $C_1'C_XC_2'/C_2'C_XC_1'$. The change in stimulus occurred when the habituation criterion was reached (2 successive drops of 20% in sucking rate). Discrimination between the two stimuli was inferred when dishabituation was clearly indicated by a strong increase in sucking rate. In the first group (CVC) 7 out of 10 subjects showed a recovery of at least 15% in the response rate, while only 2 subjects in the second group (CCC) reacted to the change in stimuli. The stimuli in this pilot experiment were direct analogic recordings of a native French speaker pronouncing the syllabic and non-syllabic sequences. In order to control overall acoustic parameters such as length and amplitude, computer normalized stimuli were used in the experiment reported here.

Method

Subjects. All of the subjects tested in this experiment were recruited at the outpatients department for newborns at the Port Royal Maternity Clinic in Paris. They were all full-term, low risk infants who weighed more than 2.5Kg at birth. 90 infants were tested, of whom only 50 were included in our experiment. Of the 40 other infants, 16 did not fulfill our habituation criterion within 12 minutes, 14 went to sleep, 7 refused to suck and 3 cried during the experiment. The 50 subjects were assigned at random to three experimental groups and one control group as follows:

1. CVC	:	N = 15;	mean age	:	37.9 days
2. CCC	:	N = 15;	mean age	:	38.6 days
3. VCCCV	:	N = 10;	mean age	:	29.2 days
4. Control	:	N = 10;	mean age	:	31.7 days

Stirhree types of stimuli (see Table 1) were used, each presented to an exntal group. All three types (CVC, CCC and VCCCV) were reconstrrom natural sounds which were sampled by the computer (PDP 1Ghigh rate (32 KHz) so as not to disturb the original acoustic quality, shaped into an analogic form. Table 1 shows the structure, the appre sound, of all the stimuli for all groups in our experiment. Of cour phonetic transcription corresponds to the way the native

TABLE 1
uation and Test Stimuli Used for the Four Groups

Grou	Stimuli				
	Habituation 'ructure	Habituation Phonetic Transcription	Duration	Test	
CVC	C_1VC_2	tap	120 msec	C_2VC_1	pat
CCC	C_1CC_2	tšp	240 msec	C_2CC_1	pšt
VCCC	VC_1CC_2V	utšpu	420 msec	VC_2CC_1V	upštu
Contr	C_1CC_2	tšp		C_1CC_2	tšp
	C_2CC_1	pšt		C_2CC_1	pšt

French stranscribed the stimuli used. For the three types of material, both thation and test stimuli were identical in length.

Ap. A blind nipple was attached to a pressure transducer which, iwas connected to an integrator. When the integrator attained a previoed threshold, by cumulating the pressure of each suck, a tape reowas switched on and the stimulus was broadcast over a speaker vicinity of the infant. A counter displayed the number of sucks mthe infant for each minute.

The three distinct phases to the experiment:

1. e: The infant's baseline sucking level without any stimulation was recorded during the first two minutes.
2. ation: During the next five minutes, the habituation stimulus was presented whenever the threshold was reached. Drops in sucking in this phase were

not considered as indicating habituation and did not trigger a change in stimulus. As of the sixth minute, however, and after two successive drops of 20% in sucking rate had been recorded, the change in stimulus took place.
3. Stimulus change: Once the stimulus had been changed, the sucking rates for the following three minutes were recorded. The experiment was terminated at the end of these three minutes. In the control group, however, where there was no stimulus change, the presentation continued for three minutes after habituation was attained.

RESULTS

Throughout this section we will use the term "dishabituation ratio" or D. R. to refer to the ratio calculated by subtracting the sucking rate before the change in stimulus from the one after the change and dividing this by the rate before the change:

$$\frac{\textit{Rate after} - \textit{Rate before}}{\textit{Rate before}}$$

Furthermore, we will consider that only infants whose D. R. is greater than 15% noticed the change in stimulus. This 15% cutoff is used to insure that the increases in sucking rate were attributable only to the infant's reaction to stimulus change. In the last minute of the habituation phase, the sucking rate of the control group does not differ significantly from that of any of the three experimental groups as shown by a student 't' test. Given that this was the case, only one control group was used throughout the experiment.

Table 2 shows the mean sucking rates for baseline and for the minute preceding and the minute following the stimulus change for all four groups. A significant difference between the sucking rates for the minute preceding and following the stimulus change was found for all experimental groups, whereas there is no significant increase in the sucking rate after habituation for the control group. Table 2 shows also the results obtained through the 't' tests carried out on these groups.

In Figure 1, we compare and contrast the dishabituation ratios for all four groups. The amount of dishabituation (Group D. R.) for the CCC group (+13.5%) is less dramatic than for the other two experimental groups, CVC (+37.2%) and VCCCV (+36.8%), and is approximately the

TABLE 2
Mean Sucking Rates for Baseline, the Minute Preceding and That Following the Stimulus Change

Groups	*Mean Sucking Rates*			*Student t (DF)*
	Baseline	*Before Change*	*After Change*	
CVC	41.9	36.5	50.1	4.82 (14)***
CCC	48.3	44.5	50.5	2.47 (14)*
VCCCV	56	45.6	62.4	3.52 (9)**
Control	46.2	37.8	44.1	1.39 (9)

*$p < .05$
**$p < .01$
***$p < .001$

same for the control group (+16.6%). As in Table 3, we divide the subjects in each group into two categories, namely, those whose D. R. was greater than 15% and those whose D. R. was less than 15%. We observe that 12 out of the 15 subjects in the CVC group show marked dishabituation, while only 6 out of the 15 subjects in the CCC group display significant dishabituation. The differences in subject distribution in CVC and CCC, and in CVC and the control group, are both significant (Fisher's Exact Probability Test, $p < .05$), whereas there is no significant difference between CCC and the control group. We note that there is a strong tendency to dishabituate for the subjects in the CVC group only. In the VCCCV group, for which we observed a high D. R. (see Figure 1), only 5 out of 10 subjects show a significant increase in their sucking rates.

In this experiment, results are presented in two ways. One refers to the change in sucking rate in the minute before and after the stimulus change.

TABLE 3
Number of Infants Who Dishabituate (D. R. > 15%), or Do Not Dishabituate (D. R. < 15%) After the Stimulus Change

Groups	*Dishabituating Infants*	*Non-Dishabituating Infants*	*Total*
CVC	12	3	15
CCC	6	9	15
VCCCV	5	5	10
Control	4	6	10

FIGURE 1

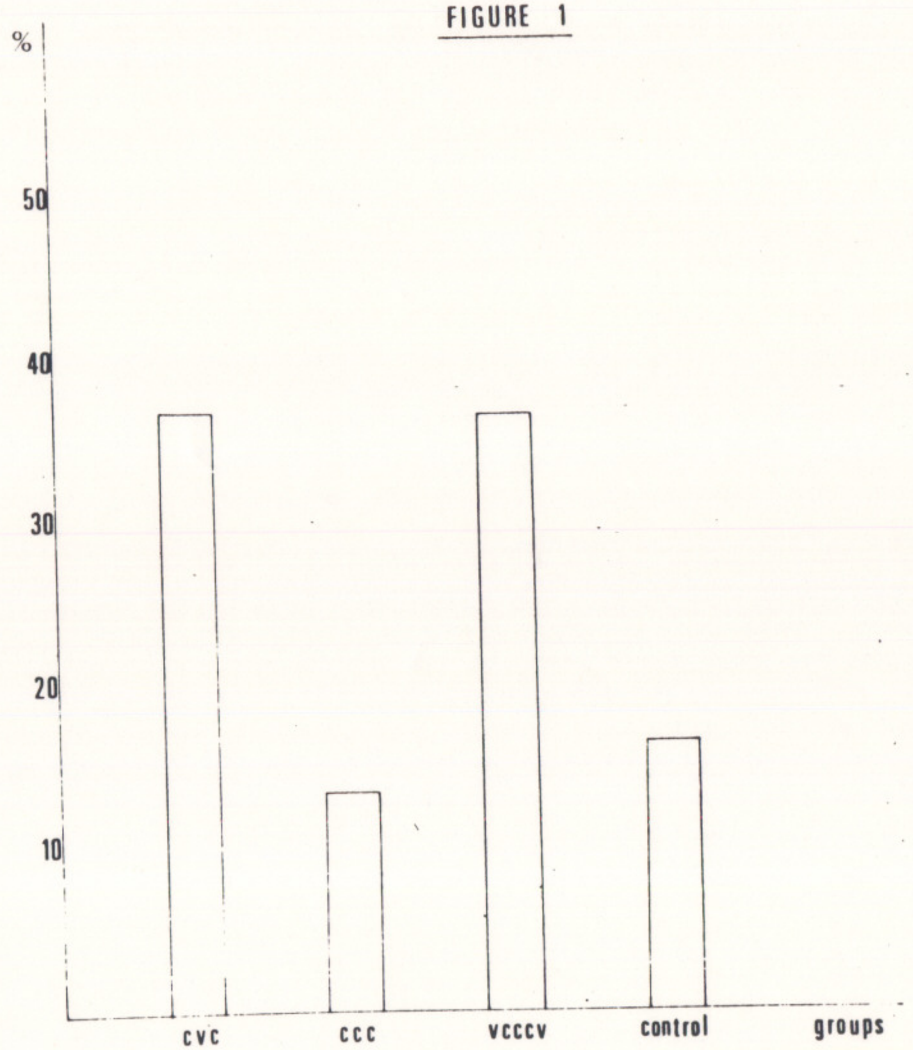

Figure 1. Dishabituation ratios for all four groups

In the other, subjects are allocated to one of two categories: one containing infants whose D. R. was greater than 15% and which we took to be those who might actually be reacting to the change, and one containing infants whose D. R. was less than 15%. Scoring all groups in terms of overall D. R. and by categorizing the number of subjects according to a D. R. of more or less than 15% has the advantage of allowing us to look at a) overall group results and b) distribution of subjects in terms of possible reaction to the change. This, in turn, allows us to make statements about the generality of certain forms of behavior. Thus, in the CVC group the D. R. is high and the number of infants who react to the change is great. Furthermore, in the CCC group, as in the control group, the D. R. is low and the number of infants who react to the change is also rather small. In contrast, the VCCCV group seems to be different in that a very high D. R. makes it look like a CVC group but a rather small number of subjects reacting to the change makes it look more like a control group, or a CCC group. This can be attributed either a very high D. R. in a few subjects or a D. R. of just under 15% in many subjects. The results show that the 5 subjects who reacted to the stimulus change had a very high D. R., while the D. R.'s of the subjects who did not react are similar in this VCCCV group to those in the other experimental groups.

DISCUSSION

In this study we explore the discrimination abilities of infants tested with well formed syllables (CVC), with impossible syllables (CCC) and with bisyllabic sequences (VCCCV) that contain as a subsegment the impossible syllable. In all the experimental stimuli the critical consonants are the same throughout, namely /p/ and /t/. For all experimental groups the difference between the habituation and the dishabituation syllables is attained in the same way: the first and the last consonants are permuted. Thus, the physical change is always the same, at least at the level of the individual phoneme. If we think of speech perception as the result of the segmentation of the speech flow into phonemes, then, indeed, the physical change in all three experiments is the same. However, under the assumption that the continuous speech flow is generally segmented into syllables, it is reasonable to predict that a change from *pat* to *tap* should result in a different acoustic representation, while a phonetically equivalent change from *pšt* to *tšp* should, generally be ignored, since neither is an object for speech processing and thus probably does not result in different acoustic representations. In particular, we predicted that *Ss* in our CVC group would discriminate much better than those in our CCC group. Furthermore, we also predicted that *Ss* in our VCCCV group would discriminate the stimuli even though the difference is phonetically identical to that in the CCC group. Nonetheless, infants may be capable of segmenting a VCCCV into two syllables that can be used for the proper representation of acoustic information in the stimuli received. Notice in this context that, even though natural languages seem to block lexical entries that are entirely constituted by three occlusive stops (or combination of non-voiced phonemes of which at least two are stops), it is

possible to find many words in natural languages with phoneme clusters between two vowels (or vowel and silence), as for instance, the word *obst* (ɔpst) in German and the word abstinence (apstinãs) in French.

As reported above, the results show that infants have no problem in discriminating two syllables that differ only in the order of the constitutive phonemes like *pat* and *tap*. Such a result was to be expected given the results reported by Eimas et al. (1971), Morse (1973) and Moffitt (1971), amongst others. Furthermore, our results indicate that infants do not seem able to discriminate two CCC's when they only differ in the order of the first and third consonants. Given that in our stimuli we used the same phonemes /p/ and /t/ throughout, it seems possible to advance that only sequences that contain an alternation of consonants and vowels (or of consonant clusters and vowels) have the structure required to be proper representations for speech. Only such representations will be preserved through time with sufficient fidelity as to permit a discriminative response to newborns.

If our hypothesis is correct, it is of some interest to understand how infants extract syllable information from the speech signal. For this purpose, we included the VCCCV group. As may be recalled, the results of that group are fairly difficult to interpret. In fact, as indicated, the DR scores and the number of *Ss* reacting to the change are not as well correlated for this group as they were for the other two. Although the number of *Ss* who increase their sucking rate by more than 15% to the change is comparable to the number in the control group, the group's DR is much higher indeed. In fact, the DR is very similar to that of the CVC group. This is due to the fact that among the 5 *Ss* who significantly increased their sucking rate after the stimulus change there are *Ss* whose reaction at that point was unduly important. We can, of course, speculate as to why VCCCV infants had difficulty with the discrimination when those who succeeded seemed to have a greater behavioral manifestation at the change, but the phenomenon is not explicable at this time. However, it may be relatively easy to understand why our *Ss* encountered some difficulty in making the VCCCV discrimination while *Ss* in Jusczyck and Thompson's (1978) experiment were so proficient at discriminating the bi-syllabic (CVCV) stimuli. First, their *Ss* were considerably older than ours. Second, there are several reasons to uphold the view that a CVCV structure may be simpler than a VCCCV. A number of linguists have tried to claim that CV is a basic underlying structure of many languages and that it may turn out to be a phonological universal. Aside from such formal considerations, our infants had to discriminate either a VC from another VC in the context of a changing CCV or, worse, discriminate a VCC from a different VCC in the context of a changing CV. The infants in Jusczyk and Thompson's experiment (1978) had to discriminate a CV from another CV in the context of an invariant nature.

The results of our experiments should be considered against the background of psycholinguistic research of the units of speech processing. As may be recalled, *Ss* who were asked to respond to a part of a word, usually presented within a list of words, performed more efficiently when the target corresponded to the first syllable of the word rather than to its first phoneme. In fact, Savin and Bever (1970) argued, on the basis of their results, that phonemes are abstract devices without either perceptual or articulatory reality. In a subsequent experiment, McNeill and Lindig (1973) systematically altered the linguistic levels of target and search lists. From their results, these investigators argue that it is the mismatch between the linguistic level of the target and that of the search list that best predicts the chronometric data obtained. These results, when considered together with those reported by Healey and Cutting (1976), may initially appear as a vindication of the primacy of the phoneme as a perceptual unit. Indeed, in McNeill and Lindig the response to a phoneme target presented in a list of phonemes yields the fastest RT of any in their experiment. However, their group that was phoneme monitoring in a phoneme search list was essentially a group that was searching for syllables in a list made up of syllabic items. Indeed, consider the way in which McNeill and Lindig constituted their material. They state that ". . . . the acoustic phoneme targets were the following plus /a/:/b/, /d/, /g/; /p/ . . . ." (pp. 421.)

Given that throughout the experiment the only vowel ever used in the phoneme list was /a/, *Ss* could easily convert phoneme targets into the corresponding CV. Subjects in the experiment reported by Healey and Cutting also encountered the same phoneme selection problem in that investigation. However, rather than using CV as phonemes, Healey and Cutting used V which in isolation have not only the status of vowel but also those of syllable or even word. These facts allow us to consider both investigations as support for the special status of the syllable as a primary device for the accessing of speech.

Using slightly different methods, some recent investigations also go in the direction in which we have been arguing. For instance, Liberman, Shankweiler, Fischer and Carter (1974) showed that small children could not play a tapping game following a phoneme segmentation rule but could master it easily when the segmentation rule was based on syllables. But the interpretation of their results remains quite elusive. Was the difficulty related to a language-specific problem, or was it that through some developmental problem or other they could master one task but not the other because of some difference in the information processing load? Morais et al. (1979) have largely contributed to solving this issue. These authors gathered data on some simple segmentation problems given to illiterate, Portuguese, adult subjects. The data gathered show that illiterates cannot delete an initial phoneme from a word nor can they add one to a word. In contrast, they can either add or delete an initial syllable to the word. Furthermore, adults of very similar background to that of the Portuguese *Ss*, but who possessed the rudiments of reading and writing in an alphabetic system, had no trouble in performing either the syllable or phoneme tasks. Though more research will be required, one interpretation that could be advanced is that the syllable is the natural speech processing unit, while the phoneme is a classificatory unit available to speakers who have mastered the alphabetic system of reading and writing. However, there is a question that comes immediately to mind. If the awareness of phonemes requires the mastering of an alphabetic system, how were such systems developed in the first place? The answer surely has something to do with the production and articulation of language. If we look at someone producing language, we can observe lip positions and even movements of the tongue. This may lead us to scrutinize our own movements, and eventually bring us to classificatory activities that give rise to phonemes or even features. However, the syllable may yet be unique as a speech processing segment.

In this experiment, it was shown that infants will discriminate a physical change better within a syllable than in any other environment. Many more results will be needed before this conclusion can be confidently generalized to all cases. We take our results to mean that infants can process units like the syllable at a very early age. We cannot make any definitive conclusion as to the comparative status of the syllable and the phoneme from the research with infants. However, future research will have to discover whether the syllable is a holistic syncretic unit which can be analyzed into phonemes or even features much later in life.

REFERENCES

Bailey, P. Perceptual adaptation in speech. Unpublished dissertation, Cambridge University, 1975.

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

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music, *Perception and psychophysics*, 1974, *16*, 564–574.

Eimas, P. D. Speech perception in early infancy. In L. Cohen & P. Salapatek (Eds.), *Infant perception: from sensation to cognition*, Volume II. New York: Academic Press, 1975.

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

Halle, M., & Stevens, K. N. Speech recognition: a model and a program for research. In J. Fodor and J. Katz (Eds.), *The structure of language*. Englewood Cliffs, N.J.: Prentice Hall, 1964, pp. 604–612.

Healey, A. F., & Cutting, J. E. Units of speech perception: phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 73-83.

Jusczyk, P. W., & Thompson, E. Perception of a phonetic contrast in multi-syllabic utterances by 2-month-old infants. *Perception and psychophysics*, 1978, *23*, 105–109.

Jakobson, R. *Six lecons sur le son et le sens*. Paris: Editions de Minuit (Arguments), 1976.

Liberman, A. M. The grammars of speech and language. *Cognitive Psychology*, 1970, *1*, 301–323.

Liberman, I. Y., Shankweiler, D., Carter, B., & Fischer, F. W. Reading and the awareness of linguistic segments. Haskins Laboratories Status Report on Speech Research, S. R. 31/32, 1972.

McNeill, D., & Lindig, K. The perceptual reality of phonemes, syllables, words and sentences. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 419–430.

Mehler, J., & Bertoncini, J. Infants' perception of speech and other acoustic stimuli. In *Psycholinguistics Series II*. John Morton and John Marshall (Eds.), Elek Scientific Books.

Menzerath, Typology of languages. *Journal of Acoustical Society of America*, 1950, *22*, 698–701.

Mermelstein, P. On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception and Psychophysics*, 1978, *23*, 331–336.

Miller, J. D., Weir, C. C., Pastore, R. E., Kelly, W. J., & Dooling, R. J. Discrimination and labelling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, *60*, 410–417.

Moffitt, A. Consonant cue perception by twenty-to-twenty-four-week-old infants. *Child Development*, 1971, *42*, 717–731.

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

Morse, P. A. The discrimination of speech and non-speech stimuli in early infancy, *Journal of Experimental Child Psychology*, 1972, *14*, 477–492.

Newman, E. B., & Gerstman, L. J. A new method for analyzing printed English. *Journal of Experimental Psychology*, 1952, *44*, 114–125.

Pisoni, D. B., & Lazarus, J. H. Categorical and non-categorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, *55*, 328–333.

Savin, H., & Bever, T. The non-perceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 295–302.

Simon, H. J., & Studdert-Kennedy, M. Selective anchoring and adaptation of phonetic and nonphonetic continua. *Journal of Acoustical Society of America*, 1978, *64*, 1338–1357.

Studdert-Kennedy, M. Status report on speech perception. Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen, August 6–11, 1979, *1*, 58–81.