Segmentation of the Spoken Word and Reading Acquisition*

Isabelle Y. Liberman+

There are many points of departure for investigators who are interested in reading. My colleagues and I at the University of Connecticut and Haskins Laboratories have begun with the fact that there are children who readily acquire the capacity to speak and listen to language but do not learn to read it. We have been concerned to ask, therefore, what is required in reading a language that is not required in speaking or listening to it.

The first answer that comes to mind, of course, is that reading requires visual identification of optical shapes. Since our concern here is with reading an alphabetic script, we may well ask whether the rapid identification of letters poses a major obstacle for children learning to read. The answer is that for most children, perception of letter shapes does not appear to be a serious problem. There is considerable agreement among investigators that by the end of the first year of school, even those children who make little further progress in learning to read generally show no significant difficulty in the visual identification of letters as such (Doehring, 1968; Kolers, 1972; Liberman, Shankweiler, Orlando, Harris, and Berti, 1971; Shankweiler, 1964; Vernon, 1960).


+University of Connecticut, Storrs. Although not a member of the Haskins Laboratories staff, the author has long been familiar with work done here. This paper reflects her interest and the Laboratories' interest in the implications of some of that work for an understanding of the reading process.

Acknowledgment: Much of the work reported in this paper was done in conjunction with Donald Shankweiler of the University of Connecticut. We are both deeply indebted to Alvin M. Liberman for many helpful suggestions. In the syllable-phoneme experiment, thanks are due to Bonnie Carter for aid in data collection and to F. William Fischer for his assistance in both data collection and statistical analysis. We are grateful also to Carol Fowler for her participation in all phases of the recent research reported here on consonants and vowels. Finally, all of us are indebted to Donald Libby, principal of Andover Elementary School, Andover, Conn., and to the teachers and pupils in that school without whose generous cooperation the research could not have been done at all.

[HASKINS LABORATORIES: Status Report on Speech Research SR-33 (1973)]
Beyond identification of letters, learning to read requires mastery of a system which maps the letters to units of speech. There is no evidence, however, that children have special difficulty in grasping the principle that letters stand for sounds. Indeed, children can generally make appropriate sounds in response to single letters, but are unable to proceed when they encounter the same letters in the context of words (Vernon, 1960).

A third possible source of difficulty is that the relation in English between spelling and language is often complex and irregular. But even when the items to be read are carefully chosen to include only those words which map the sound in a simple, consistent way and are part of the child's active vocabulary, many children continue to have difficulties (Savin, 1972).

What then are the real difficulties faced by the child in the early stages of reading acquisition? In this paper, I will explore one possible source of difficulty that has been recently proposed by us (Liberman, 1971; Shankweiler and Liberman, 1972) and other investigators (Elkonin, 1973; Klima, 1972; Mattingly, 1972)—that reading requires of the child an awareness of the structure of his language, an awareness that must be more explicit than is ever demanded in the ordinary course of listening and responding to speech. Since an alphabet is a cipher on the phonemes of a language, we should think that learning to decipher an alphabetically written word (as opposed to memorizing its visual configuration as may be done in learning so-called "sight" words) would require an ability to be quite explicit about the phonemic structure of the spoken word. For example, if the child is to map the printed word, "bat," which obviously has three letters, onto the spoken word which he already has in his lexicon, he must know that the spoken word also has three segments.

We suspect that this knowledge about the structure of the spoken word is not readily available to the child. Indeed, it appears not to have been readily available to the race. We know that an alphabetic method of writing, which rests upon an explicit phonemic analysis of the language, has been invented only once and is a comparatively recent development in the history of writing systems (Gebl, 1963). Syllabaries and logographic systems of writing, on the other hand, preceded the alphabet by thousands of years and have been invented independently several times. Of more immediate relevance to us is the evidence that children with reading disabilities often have difficulties even with spoken language when they are required to perform tasks demanding some degree of explicit segmentation of phonemic structure. These children are often reported, for example, to be deficient in rhyming, in recognizing that two different monosyllables may share the same first (or last) phonemic segment (Monroe, 1932), and, according to recent research (Savin, 1972), also in speaking Pig Latin, which demands a conscious shift of the initial phonemic segment to the final position in the word.

A third line of evidence suggesting that knowledge of phoneme structure is not readily available is provided by the behavior of reading-disabled children as observed by teachers who have worked with them (Johnson and Myklebust, 1967). Such a child will often demonstrate, as I have suggested earlier, that he can readily recover the phonemic segments in the ordinary course of speaking and listening. That is, he can respond appropriately to spoken words and to the objects to which they refer. Moreover, he can approximate the letter-to-sound correspondences. If he is asked, for example, to give the sound of the letter "b" he will say /bA/. For the sound for the letter "a" he will say /ae/ (though
this may give him more trouble, as we will discuss later). For the sound of the letter "t" he will say /tʌ/. But then if he is shown the printed word "bat" and asked to read it, he may give any one of a variety of incorrect responses (which I will deal with in more detail below in a discussion of error analysis). If he is then pressed to try to "sound it out," or otherwise to use what he knows about the letter-to-sound correspondences, he is likely to produce /bʌ/ /ae/ /tʌ/. At that point, he may be urged by the teacher to "say it faster," "put the sounds together," or, in the phrase commonly used, to "blend it." But no matter how fast he produces those sounds or how desperately he tries to put them together, he produces a nonsense word "buhatuh" containing five phonemic segments and not the word "bat," which has only three. Somehow, he cannot relate the three letters of the printed word to the three phonemic segments of the spoken word. It is as if he were not aware of the fact that the monosyllabic spoken word has three segments.

But why should it be so difficult for the child to become explicitly aware of phonemic segmentation? If, as has often been supposed, the sounds of speech bore a simple one-to-one relation to the phonemic structure just as the letters do (at least in the orthographically regular case), it would indeed be hard to see why phonemic analysis should pose special problems. That is, if there were in the word "bat" three acoustic segments, one for each of the three phonemes, then the segmentation of the word that is represented in its spelling would presumably be readily apparent.

However, as extensive research in speech perception has shown (Fant, 1962; A. Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Stevens, 1972), the segmentation of the acoustic signal does not correspond directly or in any easily determined way to the segmentation at the phonemic level. Moreover, this lack of correspondence does not arise because the sounds of the phonemes are merely linked together, as are the letters of the alphabet in cursive writing or as may be implied by the reading teacher who urges the child to blend "buhaguh" into a word that he knows. Instead, the phonemic segments are encoded at the acoustic level into essentially unitary sounds of approximately syllabic dimensions. In the case of "bat," for example, the initial and final consonants are folded into the medial vowel, with the result that information about successive segments is transmitted more or less simultaneously on the same parts of the sound (A. Liberman, 1970). In exactly that sense, the syllable "bat," which has three phonemic segments, has but one acoustic segment.

This is not to say that the phonemic elements are not real, but only that the relation between them and the sound is that of a very complex code, not a simple, one-to-one substitution cipher (A. Liberman et al., 1967). To recover the phonemic segments, to sort them out from the complex code, requires a correspondingly complex decoding process. In the normal course of perceiving speech these processes go on tacitly and automatically. To understand speech the listener need not be any more aware of the phonemic structure than he is of the rules of syntax.

Since the acoustic unit into which the phonemic elements are encoded is of approximately syllabic dimensions, one might suppose that the number of syllables (though not necessarily the exact location of the syllable boundaries) would be more readily apprehended than the phonemes. Syllable segmentation may be easier than phoneme segmentation for another reason as well. There are peaks of acoustic
energy (hence loudness) that correspond at least roughly to the vocalic nucleus of the syllable (Fletcher, 1929). Thus the syllable is acoustically marked, while the phoneme is not.

If syllabic segmentation is indeed easier, we might then have an explanation for the assertion (Makita, 1968) that the Japanese kana is readily mastered. The kana, one of the two Japanese writing systems, is approximately a syllabary. That is, most of the graphic symbols in the kana represent a syllable rather than a phoneme. There are separate symbols for ba, be, bi, bu, ga, ge, gi, gu, etc. Given the open syllable (CV) structure of the Japanese spoken language, the child therefore rarely needs to go below the level of the syllable in order to master the writing system. One might expect, further, that an orthography which represents each word with a different character (as is the case in Chinese ideographs or in the closely related Japanese kanji) would also not cause, in the beginning reader at least, the particular difficulties that arise in mastering the more analytic alphabetic system. Indirect evidence of the special burden imposed on the beginning reader by an alphabetic script can be found in the relative ease with which reading-disabled children learn kanji-like representations of language while being unable to break the alphabetic cipher (Rozin, Poritsky, and Sotsky, 1971). It is worth noting, in addition, that since the time of the ancient Greeks, methods of reading instruction have sporadically reflected the assumption on the part of educators that the phonemic structure of the language is more easily taught through the initial use of syllabic units (Mathews, 1966).

Though these considerations are suggestive, there has been no direct empirical test of the assumption that young children do, in fact, find it difficult to make an explicit phonemic analysis of the spoken word and that this ability comes later and is more difficult than syllabic analysis. My colleagues and I have undertaken in a recent experiment to provide such a test. For that purpose, we asked how well children can identify the number of phonemic segments in spoken words and how this compares with their ability to deal similarly with syllables.

The subjects were four-, five-, and six-year-olds in preschool, kindergarten, and first grade classes, respectively. They included 46 preschoolers, 49 kindergarteners, and 40 first graders. The unequal numbers arose from our plan to include all available children in the particular school at each grade level. Alphabetized class registers were used at each grade level to divide the children into the two experimental groups, one assigned to phoneme segmentation and the other to syllable segmentation. The level of intelligence of all the subjects was roughly assessed by means of the Goodenough Draw-a-Person Test. Two-way analyses of variance performed on the Goodenough DAP scores revealed no significant differences in IQ, either across tasks or across grade levels. The mean chronological ages of the two task groups were also not significantly different. Therefore, any performance differences in the two types of segmentation can reasonably be taken to reflect differences in the difficulty of the two tasks.

The procedure was in the form of a tapping game. The child was required to repeat a word or sound spoken by the examiner and to indicate, by tapping a wooden dowel on the table, the number (from one to three) of the segments (phonemes in one group, syllables in the other) in the stimulus items. Four sets of training trials containing three items each were given to both groups. The test trials, which followed the four sets of training trials, consisted of 42 randomly assorted individual items of one, two, and three segments which were
presented without prior demonstration and corrected, as needed, immediately after the child's response. Testing was continued through all 42 items or until the child reached criterion of tapping six consecutive trials correctly without demonstration. Instructions given to the two experimental groups at all three age levels were identical except that the training and test items involved phonemic segmentation in one group and syllabic segmentation in the other. All the children were tested close to the end of the school year.

The results showed in many ways that the test items were more readily segmented into syllables than into phonemes. In the first place, the number of children who were able to reach criterion was markedly greater in the syllable group than in the phoneme group, whatever the grade level. At age four, none of the children could segment by phonemes, while nearly half could meet the stringent criterion with the syllables. Ability to perform phoneme segmentation successfully did not appear at all until age five, and then it was demonstrated by only 17% of the children. In contrast, almost half of the children at that age could segment syllabically. Even at age six, only 70% succeeded in phoneme segmentation, while 90% were successful in the syllable task.

The contrast in difficulty can also be seen in terms of the number of children who achieved criterion level in six trials, which, under the procedures of the experiment, was the minimum number possible. For the children who worked at the syllable tasks, the percentage reaching criterion in the minimum time increased steadily over the three age levels. It was 7% at age four, 16% at age five, and 50% at age six. In striking contrast to this, we find that in the phoneme group, no child at any grade level attained the criterion in the minimum time. An analysis of variance which assessed the contribution of task and grade found that these main effects were highly significant, with a p level of less than .001.

We cannot judge from this experiment to what degree the measured increases in phoneme segmentation with age represent maturational changes and to what extent they may reflect the effects of instruction in reading. We would guess that the sharp increase from 17% at age five to 70% at age six in the number of children passing the phoneme task is probably due in large part to the intensive concentration on reading and readiness activities in the first grade. The possibility that these changes with age between five and six are relatively independent of instruction could be tested by a developmental study in a language community such as the Chinese, where the orthographic unit is the word and where reading instruction therefore does not demand the kind of phonemic analysis needed in an alphabetic system.

Meanwhile, we are especially concerned to know more about those substantial numbers of first graders, some 30% in our sample, who apparently have not acquired the ability to do phoneme segmentation. It would be of primary interest to know whether they will show deficiencies in reading acquisition as well. We are just beginning this phase of the research. In a recent pilot study, we gave the word-recognition subtest of the Wide Range Achievement Test (the WRAT) to the children who were the first graders of last June's sample. When they are ranked according to their scores on the reading test, we find that while half the children in the lowest third of the class in reading ability had failed the phoneme segmentation test last June, no child in the top third had failed it. Encouraged by these results, we have devised an analytic reading test designed to measure
decoding skills more systematically than is possible with the WRAT. This is now being administered in addition to the WRAT and the phoneme task to a new group of children in Grades 1 and 2.

We have suggested that a lack of awareness of phonemic segmentation may be one serious roadblock to reading acquisition. Data from the analysis of children's reading errors appear to provide additional indirect evidence for this view. It seemed to us that if a child's chief problem in reading is that he cannot make explicit the sound structure of the language, he might be expected to show success with the initial letter which requires no further analysis of the syllable and to show relatively poor performance beyond that point. If all he knows are the letter-to-sound correspondences and that he must proceed from left to right, he might in the case of "bat," for example, simply pronounce the sound for the first letter and then search his lexicon for a word beginning with the sound of that letter. What he needs to do, instead, is to search his lexicon for a word that has three sound segments corresponding to the letter segments in the printed word. However, if he does not know that the words in his lexicon have segments or if he finds phonemic segmentation difficult, he will not be able to map the letters to the segments in those words. By this reasoning, his errors on the final consonants in words should be greater than those on the initial consonants.

We have recently concluded an experiment designed specifically to examine the initial-final consonant error pattern. The subjects were 20 third graders drawn consecutively from the alphabetic registers of a nearby elementary school. The list of words to be read consisted of 38 monosyllables familiar to third graders and selected so as to give equal representation to the 19 consonant phonemes which can occur in both initial and final position in English words. Each phoneme was represented twice in the list in each position. The words were printed on 3 x 5 cards and presented to the child singly to be read aloud to the best of his ability. Testing was carried out in late fall.

Analysis of the data shows final consonant errors to be about twice as frequent as initial (9.5% of the opportunities for final consonants as compared with 4.9% for those in the initial position). A t-test found this difference to be highly significant, with a p value of less than .005. Since it was possible that the difference might be due to the fact that a given phoneme occurring finally may be spelled more complexly than that same phoneme in the initial position (g, j versus dge or ge), we then looked only at the errors on phonemes which are spelled simply (by a single letter) in both initial and final position (p, t, k, b, d, g, m, n, r). If the difference had been due to orthographic complexity, it should have disappeared in this analysis. But it did not. Final consonants still produced significantly more errors (7.8% to 3.0%).

It is clear, then, from these results, that there is indeed a progression of difficulty with the position of the segment in the word, the final consonants being more frequently misread than the initial. Similar findings have been reported by us in a previous study using different word lists (Shankweiler and Liberman, 1972) and by other investigators (Daniels and Diack, 1956; Weber, 1970) who examined error patterns in the reading of connected text. This initial-final consonant difference cannot be accounted for in terms of a simple reflection of the error pattern in speech, as we found in the earlier study of error patterns. There we presented, first for oral repetition and then for reading, a list of
204 monosyllables chosen to give equal representation to most of the consonants, consonant clusters, and vowels of English. The initial-final consonant error pattern was duplicated in reading, but in oral repetition, the consonant errors were about equally distributed between initial and final position. Moreover, the initial-final error pattern in reading is also contrary to what would be expected in terms of sequential probabilities. If the child at the early stages of beginning to read were using the constraints built into the language, he would make fewer errors at the end than at the beginning of words, not more.

Thus far we have presented several lines of evidence suggesting that the explicit analysis of phoneme segmentation is a hard and unnatural task which may be an important source of difficulty for the child learning to read. But it is certainly not the only serious barrier. The error pattern of vowels provides a case in point. It is well established (Monroe, 1932; Shankweiler and Liberman, 1972; Venezky, 1968; Weber, 1970) that vowels elicit many more errors than consonants. In the segmentation study mentioned above, for example, the vowel errors were twice as frequent as overall consonant errors (15.1% for the vowels as compared with 7.3% for the consonants). It should be noted that this is quite different from what we find in speech. The vowel errors in the oral repetition of speech are infrequent and fewer than those for consonants (Shankweiler and Liberman, 1972).

Why should the error rate for reading vowels be so much higher than that for consonants? It might, of course, be simply because of the embedded medial position of the vowel in the words used to test reading. In order to check on this possibility, we devised a new test consisting of equal numbers of words containing vowels in the initial, medial, and final positions. The seven vowel phonemes that can occur in all three positions were used three times in each position. The words were again monosyllables familiar to third graders. It was found that the overall rate of vowel errors continued to be about twice that of consonant errors (28.3 to 14.0).

There are two reasons at least for suspecting that vowel errors may reflect something other than the segmentation problems which we have suggested as an explanation for the consonant pattern. First, as we have seen, the child can apparently count syllables fairly well and the vowel nucleus stands out in the spoken word as a major element that can be identified in the syllable. A second, and perhaps more interesting reason, comes from a further examination of the error pattern. In the case of consonants, we have noted that errors tend to pile up in the final position. We have taken this as indirect evidence that the child is having segmentation problems. Vowel errors, on the other hand, pattern quite differently. In the third grade study described above, there was no significant difference in error rate for vowels in the initial, medial, and final positions. Moreover, the error rate of vowels in both initial and final position continued to be significantly higher as compared with consonant errors in the corresponding positions (27.6% to 9% in the initial position and 30.5% to 19% in the final position).

Clearly there is no position effect with the vowels; they are simply difficult in all positions. The absence of a position effect may be due to the fact that the vowel is acoustically marked by a burst of sound wherever it appears, while there is no such acoustic mark for the enfolded consonant. In any event, the vowel problem certainly cannot be attributed entirely to segmentation difficulties.
Indeed, we suspect that the errors elicited by consonants and vowels are quite different in their origins. In the case of the consonants, the child has little trouble in learning the spelling-to-sound correspondences. Orthographic complexity makes no appreciable difference to the position effect. The child's error pattern arises mainly from the fact that he cannot map the segmentation of the printed word to the segmentation of the spoken word. The extra difficulties attendant upon the vowels are probably due in part to the obvious orthographic complexities of the spelling-to-sound correspondences, but partly also to the continuous and fluid nature of vowel perception (A. Liberman et al., 1967; A. Liberman, 1970). Though it stands out wherever it occurs in speech, the vowel is complicated by the fact that it can be spelled in many ways in the writing system and is less categorically perceived than the consonants. That is, not only is there a many-to-one mapping of spelling to sound, but because of the continuous nature of vowel perception, even the sound correspondences of single vowel letters may be harder to code and to maintain in memory. We have argued (Shankweiler and Liberman, 1972) that as a consequence of the continuous nature of their perception, vowels tend to be somewhat indefinite as phonologic entities, as illustrated by the major part they play in the variation among dialects and the persistence of allophones within the same geographic locality. By this reasoning, it could be that the noncategorical nature of vowel perception may itself be one cause of the complex orthography and at least one reason why multiple representations of the vowels are tolerated.

The investigation of the effect of orthographic complexity is beset with many problems. To cite only one example: if orthographic complexity is an important source of errors, the number of possible orthographic representations of a given sound should be correlated with the number of errors made on that sound. In fact, however, in a group of second graders we studied recently, the correlation between orthographic complexity and the number of errors lacked statistical significance. Qualitative analysis of the data suggests that this might be due not to the unimportance of orthographic complexity, but rather to the fact that the second grader's knowledge of orthographic rules is so slight that the number of orthographic representations is not yet a relevant factor in determining his errors. We have since developed a cloze procedure test to measure knowledge of orthographic rules for checking our findings, but these data are not yet completed.

Though we believe it is of interest to examine the relation of orthographic complexity of the vowels to the problems of reading acquisition, we recognize that the vowel may be less important in the process than would first appear. It could be argued that if the child's segmentation problems were corrected, his difficulties with the vowels would not be such a serious barrier to reading acquisition. The consonants carry most of the information load. Provided the child knew how many consonants there were and their sequence in the spoken word, an incorrect rendition of the vowel sound would be fairly easily corrected in the context. Surely, getting the vowel correct without a proper analysis of the phonemic structural sequence of the word would be of less benefit to him. If this is so, early teaching methods which emphasize the intensive teaching of the phonemic structure of the word before the introduction of letter forms should be considered. Elkonin (1973), a Russian psychologist, has recently presented considerable experimental evidence that such a method is indeed highly successful.
REFERENCES


165
