JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 20, 298–305 (1981)

# The Syllable's Role in Speech Segmentation

JACQUES MEHLER, JEAN YVES DOMMERGUES, AND ULI FRAUENFELDER

*CNRS*

AND

JUAN SEGUI

*Université René Descartes et EPHE associé au CNRS*

In this study a monitoring technique was employed to examine the role of the syllable in the perceptual segmentation of words. Pairs of words sharing the first three phonemes but having different syllabic structure (for instance, *pa-lace* and *pal-mier*) were used. The targets were the sequences composed of either the first two or three phonemes of the word (for instance, *pa* and *pal*). The results showed that reaction times to targets which correspond to the first syllable of the word were faster than those that did not, independently of the target size. In a second experiment, two target types, V and VC (for instance, *a* and *al* in the two target words above) were used with the same experimental list as in experiment one. Subjects detected the VC target type faster when it belonged to the first syllable than when it belonged to the first two syllables. No differences were observed for the V target type which was in the first syllable in both cases. On the basis of the reported results an interpretation in which the syllable is considered a processing unit in speech perception is advanced.

The perceptual units into which the continuous speech signal is segmented constitute one of the central issues in psycholinguistics. Investigators have argued for the perceptual primacy of a wide variety of candidates such as the phoneme, syllable, word, and sentoid. Traditionally, the phoneme has been taken as the basic perceptual unit. However, the lack of invariance between the acoustic signal and the phoneme has weakened its candidacy. Indeed, the acoustic information corresponding to a phoneme is often distributed throughout several segments or inversely, the information about several phonemes is concentrated into one short stretch of acoustic signal (Liberman, 1970).

Such facts as these have led some to abandon the phoneme and endorse the syllable as the basic segmentation unit. Accordingly, some device might operate automatically on the speech wave segmenting it into syllabic units. The feasibility of such a view is corroborated by the frequent incorporation of the syllable into speech recognition systems (Mermelstein, 1975; Vaissière, 1980). Further evidence in support of the syllable has also accumulated in other domains. By means of a recognition masking technique, Massaro (1974) has shown the existence of a 250-millisecond processing and storage period which defines a unit corresponding to the syllable.

There are also some data bearing on the classificatory status of the syllable in young children and adults. Liberman, Shankweiler, Fischer, and Carter (1974) have tested young children on a tapping game. They showed that children (4–5 years old) could identify by the number of taps syllabic, but not phonetic, segments; only older children (6–7 years old) were able to perform on both tasks. Inspired by

The authors wish to express their thanks to Yves Barbin for his invaluable technical assistance. This work was supported by grants from Harry Frank Guggenheim Foundation 1977/1978 and D.G.R.S.T. 1980 to J. Mehler. Requests for reprints should be sent to Dr. J. Mehler, Laboratoire de Psychologie, CNRS, 54, Boulevard Raspail, 75006 Paris.



0022-5371/81/030298-08$02.00/0

these results, Morais, Cary, Alegria, and Bertelson (1979) explored whether this performance difference is due to cognitive growth or to some explicit training procedure (learning to read and write usually occurs between ages 4 and 7). They have shown that illiterate adults perform more or less like the younger children. Neither can add phonemes to or delete them from words or nonwords. However, adults who had only recently learned the phoneme–grapheme correspondence system were able to perform on both phoneme and syllable tasks. These results suggest that the syllable is a unit available for classificatory purposes early, whereas the phoneme is only available to those having mastered the phoneme–grapheme correspondence.

Some more direct evidence concerning the role of the syllable in speech recognition comes from monitoring studies. In a study conducted to test the perceptual reality of phonemes and syllables (Savin & Bever, 1970), subjects were asked to detect as quickly as possible either the first phoneme of a syllable or the syllable itself presented in a list of nonsense syllables. The systematically shorter RTs obtained for syllables led Savin and Bever to conclude that phonemes are not perceived directly but are derived from an analysis of the syllabic perceptual unit.

Both the methodology and the interpretation in this experiment have come under a certain amount of attack. Foss and Swinney (1973) propose an alternative explanation in which an important distinction is made between perception and identification. They maintain that, generally, "smaller units are identified by fractionating larger ones" (p. 254). This hypothesis was extended to predict that the detection of phonemes and syllables implies the identification of the word in which they are found.

Initial empirical support for this prediction was provided in a study by Rubin, Turvey, and Van Gelder (1976) who found that initial stop consonants were detected faster in words than in nonwords. However, these results were later attributed to the use of a modified monitoring procedure in which subjects had to monitor two targets simultaneously. The abnormally long RTs obtained with this procedure suggest that it induced a lexically based response. More recent research has shown that the detection of item initial phonemes and syllables can be based on a prelexical "phonetic" code (Foss & Blank, 1980; Segui, Frauenfelder, & Mehler, note 1) suggesting that the perception or identification of a linguistic unit at level $n$ is not necessarily derived from a higher level $n + 1$.

Many of the other studies, critical of Savin and Bever (McNeill & Lindig, 1973; Healy & Cutting, 1976; Swinney & Prather, 1980; Mills, 1980) suffer from some serious methodological weaknesses (see Mehler, Segui, & Frauenfelder, 1980 for more details). Thus, for example, the phoneme conditions used in an alleged comparison between phoneme and syllable detection times (Swinney & Prather: one vowel condition; Mills: a match condition) are most likely syllable conditions. Finally, none of these alternative explanations can explain the strong correlation between RTs to phonemes and to the syllable in which they are found (Segui et al., note 1); a result consistent with the original hypothesis proposed by Savin and Bever.

To examine the role of the syllable in speech segmentation further, RTs to sequences of phonemes which did or did not correspond to the syllabic structure of the stimulus words were compared. Thus, for example, although the words *palace* and *palmier* share the first phonemes /p/ /a/ /l/, they have different syllabic structures: *pa* . . . and *pal* . . . , respectively. Subjects were given target phoneme sequences (*pa* and *pal*) to detect in stimulus words of both syllabic structures. If the stimulus words are segmented according to their syllabic structure, RTs should be faster when the target phoneme sequence matches the first segmented syllable of the stimulus word. On the other hand, two alternative pho-

neme based predictions can be made by hypothesizing a mechanism not based on syllabic structure but on the size of the target. First, if the subject conducts a phoneme-by-phoneme analysis of the stimulus, shorter RTs could be expected for the shorter target sequences (RT: *pa* < *pal*) since less of the stimulus word (one fewer phoneme) must be analyzed to initiate a response. Finally the inverse results (RT: *pa* > *pal*) is predicted by the uncertainty hypothesis (Foss & Swinney, 1973; Swinney & Prather, 1980) according to which RTs depend on the uncertainty of the subjects concerning the nature of the target/stimulus. The smaller the target (in number of phonemes?), the fewer cues there are available for identifying the stimulus and hence the longer the RTs.

These three hypotheses will be put to test in the following experiment.

## Experiment I

### *Method*

*Subjects.* Forty-two subjects (divided into two groups of 21 each), all native French speakers from the Parisian university community, participated in the experiment which lasted about one half hour.

*Materials and design.* Five pairs of monomorphemic bisyllabic French nouns of similar frequency sharing the same initial three phonemes (CVC) were selected such that these phonemes made up the first syllable for one member of the pair and the first two (CV) phonemes formed the first syllable of the second member of the pair. For instance, in the pair: *palace/palmier* the first three phonemes (/p/ /a/ /l/) are identical. Yet, this CVC sequence corresponds to the first syllable only the word *palmier*, CV being the first syllable of the word *palace*. For each of the five pairs the initial consonant was either a voiced or a voiceless stop and the second was a liquid (either /l/ or /r/). The five pairs were: *palace–palmier, carotte–carton, tarif–tartine, garage–gardien, balance–balcon.*

These 10 stimulus words containing the target were each placed as the last item in experimental sequences. Each experimental sequence was composed of the target item and 1 to 4 bisyllabic filler words. Both members of a pair appeared in the same position (from the second to the fifth position) in their respective sequences. In addition to this first set of 10 target sequences, a second set was used with different fillers but with the same target words in the same position as in the first set. Each set of target sequences was mixed with 10 different distractor sequences forming two blocks (A & B), each containing 20 sequences. Distractor sequences either had target words as the last item (any position from the first to the sixth) or no target to prevent the subjects from anticipating last items in long sequences. These 40 sequences along with 5 warm-ups were recorded by a French native speaker at a normal rate with a two-track Ampex AG440B. The words in each sequence were separated by 2-second intervals; sequences were separated by 10 seconds.

Table 1 provides a description of the experimental sequences and of the targets given to the two groups of subjects. Two groups were used to counterbalance the presentation order of a given stimulus word and its two corresponding targets.

As can be seen in Table 1, the stimulus word *balance* in position 2 is associated with the target /bal/ for group 1 and with /ba/ for group 2 in sequence number 6 of Block A. The inverse matching is found in sequence number 19 of Block B. The targets to be detected were displayed visually on small 3 × 5 inch cards numbered from 1 to 45. Before hearing each sequence, subjects heard the instruction "next card" and had 10 seconds to turn and read the next card. Subjects in groups 1 and 2 received different decks of cards.

The list was presented binaurally and subjects responded by pressing a response button that stopped an electronic clock in a PDP-12 computer. A click, aligned manu-

TABLE 1
DESCRIPTION OF EXPERIMENTAL SEQUENCES AND TARGETS

| Number of experimental sequence | The position of the stimulus words in the experimental sequences | | | | Target for subjects in Group 1 | Target for subjects in Group 2 |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | |
| Block A | | | | | | |
| 6 | MORCEAU | *BALANCE* | | | BAL | BA |
| 8 | JEUDI | MUSEE | *CAROTTE* | | CA | CAR |
| 17 | FICHIER | *BALLON* | | | BA | BAL |
| 25 | SERVICE | CHEVEUX | *CARTON* | | CAR | CA |
| Block B | | | | | | |
| 19 | MONSIEUR | *BALANCE* | | | BA | BAL |
| 33 | EPOQUE | MAISON | *CAROTTE* | | CAR | CA |
| 40 | PROBLEME | CHEVAL | *CARTON* | | CA | CAR |
| 42 | JOURNEE | *BALCON* | | | BAL | BA |

ally with the beginning of the target word, triggered the clock. A correction for each click was obtained by means of a two channel oscilloscope and was edited into the data collection program.

## RESULTS

The mean reaction times for each subject and for each item were computed. RTs longer than 1000 milliseconds and shorter than 100 milliseconds were omitted from the calculation of the means. The omitted data made up less than 3% of the subjects' responses.

The results obtained in this experiment are displayed in Figure 1. These results show that when the subjects have to respond to a CV target in a word whose syllable structure is CV/ . . . (hereafter referred to as CV words) their RTs are faster than when they respond to the same word with a CVC target (352 msec vs 371 msec, respec-

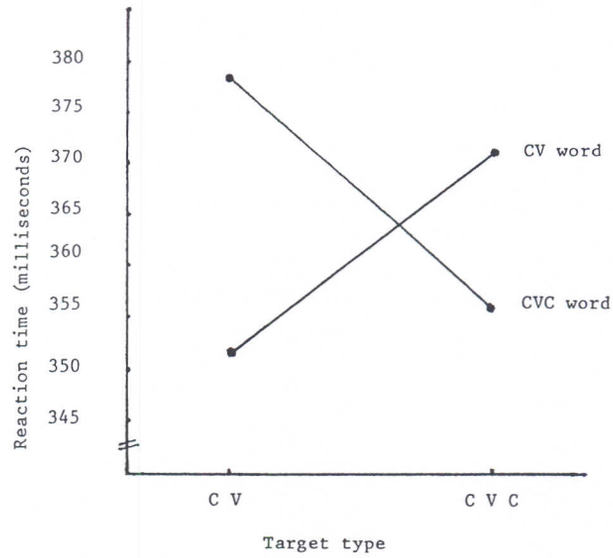


FIG. 1. Mean reaction time in milliseconds for CV and CVC words as a function of target type.

tively). When responding to a CVC word with a CVC target, subjects respond faster than when responding to the same word with a CV target (356 msec vs 378 msec, respectively). Since the results for Group 1 and Group 2 are analogous, a two-way within-subject analysis of variance was conducted using the combined data. The two main factors were target type (CV vs CVC) and word type (CV words vs CVC words). The interaction between these factors is highly significant, $F(1,42) = 11.03, p < .005$, but individually they failed to yield any significant difference, $F < 1$ in both cases. Likewise, a two-way analysis of variance was carried out on the mean reaction times corresponding to the five pairs of words giving $F(1,4) = 6.75, p < .10$ for the interaction between the two main factors. As before, individually these factors were not significant.

A $t$ test was used to compare the mean RTs for both word types as a function of the target type. Using the subjects as a random variable, a significant difference was obtained for both word types; for CV words, $t(41) = 2.02, p < .02$ (one tailed), for CVC words, $t(41) = 2.59, p < .01$. A $t$ test using items as the random variable gave analogous differences for CV words, $t(4) = 2.67, p < .05$, for CVC words, $t(4) = 2.36, p < .05$.

## Discussion

These results show that subjects detect a target phoneme sequence faster when it corresponds to the first syllable of the stimulus word than when it does not. Thus, the RTs do not depend on target size as predicted by the two alternative hypotheses but rather on the syllabic relationship between the target and the stimulus word. The following (albeit simplistic and schematic) description of what subjects are likely to do in this task provides insight into these results. First, upon seeing the target, subjects must form and store some representation of it. On subsequently hearing the acoustic stimulus, they must segment and analyze it to produce a stimulus representation. The detection process can thus be seen as the matching of stimulus and target representations.

The analysis of the acoustic signal corresponding to a given target word (*palmier*) is unlikely to vary as a function of the target (/*pa*/, /*pal*/) given to the subject. The observed RT differences can then be attributed to the process of matching between the stimulus and target representations and not to the computation of the former. When the target sequence corresponds exactly to the first syllable of the stimulus word, there is a better match between the two representations than when it does not. The varying degree of compatibility between the target and the stimulus representation could thus explain the direction of the RT differences. Since the representation of a given target (*pa*) presumably does not vary, the processing of the two words in a pair (*palmier, palace*) must differ. Thus there must be acoustic cues in the signal that the subjects exploit to derive different stimulus representations for the two words. This interpretation can explain the highly significant interaction found between word and target types.

This account is compatible with the target–stimulus mismatch hypothesis proposed by Mills (1980a; 1980b). Mills argues that the closer the match is between the subjects' expectancy about the stimulus and the actual acoustic stimulus, the faster the subjects respond to the stimulus. It should be noted, however, that this explanation can not account for the RT differences found here without appealing to a syllabic segmentation unit. Indeed, a target–stimulus mismatch hypothesis, formulated in phonemic terms would not predict the observed interaction. If the subjects' uncertainty concerning the stimulus is based only on the number of phonemes in the target, then longer targets should have led to shorter RTs. Thus, a syllable-based hypothesis is necessary. Unfortunately, however, our experiment does not allow us to make precise claims

concerning the target representation and the more-or-less analyzed character of the syllable.

In the hope of better understanding the role of the syllable and its potential constituents in word processing, a second experiment was carried out. This experiment was based on the assumption that if subjects segment the signal syllabically, then they would respond faster to a sequence as /al/ when it is contained in the same syllable *palmier* than when it is found in two different syllables *palace*. This prediction is derived from the general hypothesis according to which subjects respond faster to stimulus items when these belong to the same rather than to different constituents at any linguistic level (Fodor, Bever, & Garrett, 1974).

## Experiment II

### Method

*Subjects.* Twenty subjects from the Parisian university community participated in this experiment.

*Materials and design.* The experimental design and the linguistic materials were the same as those used in the first experiment. Two groups of 10 subjects received the same instructions with the exception that they were told that the visually specified targets V and VC could appear anywhere in the word and not just in the initial position as in the first experiment. For the experimental words, the V target corresponded to the vowel in the first syllable and the VC target to the first VC sequence in the word.

### Results

The mean reaction times for each subject and for each item were computed. Reaction times longer than 1500 milliseconds and shorter than 100 milliseconds were omitted from these calculations. The excluded data made up 6.5% of the subjects' responses. Table 2 shows the overall reaction times.

Table 2 shows that subjects' RTs to VC targets in CV words are significantly slower than those to the same target in CVC words, $t(9) = 1.76$, $p < .05$ (one tailed). On the other hand, no differences were found between RTs to V targets in CV and CVC words.

TABLE 2
Mean Reaction Times (msec) for CV and CVC Words as a Function of Target Type

| Target type | CV word | CVC word |
|---|---|---|
| V | 615 | 614 |
| VC | 704 | 658 |

### Discussion

The results observed for VC targets support our hypothesis, according to which faster RTs are expected when the target sequence belongs to one syllable rather than to two different syllables. Furthermore, no differences were obtained for the V target in a CV or a CVC syllable. These two results taken together are compatible with the syllabic segmentation hypothesis. However, the longer overall RTs of the order of 650 milliseconds do not eliminate the possibility of a postlexical access response. An experiment by Marslen-Wilson (note 2) and some preliminary results obtained in our laboratory suggest that the task of monitoring targets anywhere in a word may lead the subject to rely more heavily on the postaccess code. If this is the case, these results may be taken as an indication that the post lexical code (phonological code) is also syllabic in structure.

## General Discussion

Results of experiments I and II are compatible with the general hypothesis that the syllable constitutes a unit of speech processing. Indeed, through a process of segmentation, subjects seem able to attain syllable-like constituents for the processing of words and sentences. However, one of the major problems with a syllabic parser was raised by Liberman and Studdert-Kennedy (1978). These authors claim that "syllable boundaries in fluent speech are

frequently random with respect to words or morphemes'' (p. 173). They challenged proponents of syllabic segmentation to account for the way in which the sentence ''*He's a repeated offender*'' is parsed syllabically in a way compatible with its morphological structure. A partial solution to this challenge may come from the signal itself and/or from top-down constraints. Recent results reported by Mills (1980) suggest that word boundary information is coded in the syllable. Mills showed that a CVC target sequence which corresponds to a word (*can*) was detected faster in the same monosyllabic word than either in this word spliced out of bisyllabic and trisyllabic words (*candle, candlelight*) or in these words themselves. Thus the word boundary cues in syllables would improve the mapping of syllable on the morphemes and words. Furthermore, top-down constraints may also help with this mapping process.

The evidence for syllabic segmentation presented here clearly has its implications for lexical access. As was pointed out earlier, the subjects' detection response probably precedes lexical access and thus is based on the prelexical code (Foss & Blank, 1980). Accordingly, the postulated syllabic segments could well serve as accessing units. This claim is at odds with lexical access models based on phonetic units. For these models, the words *palmier* and *palace* are still potential word candidates (in the same cohort) for the subject having heard the sequence /*pal*/. In the syllabic hypothesis, these two words could be distinguished earlier because their syllabic structure furnishes more information than was usually assumed.

The results reported here provide evidence for syllabic segmentation of speech. However, more research is necessary to determine which acoustic cues are actually used in syllabic segmentation and also in lexical access.

## REFERENCES


FODOR, J. A., BEVER, T. G., & GARRETT, M. F. *The psychology of language*. New York: McGraw–Hill, 1974.

FOSS, D. S., & BLANK, M. A. Identifying the speech codes. *Cognitive Psychology*, 1980, **12**, 1–31.

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

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

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

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. Phonetic perception. In R. Held, H. W. Leibowicz, & H. L. Teuber (Eds.), *Handbook of sensory physiology*. VII. *Perception*. Berlin: Springer-Verlag, 1978.

LIBERMAN, I. Y., SHANKWEILER, D., FISCHER, F. W., & CARTER, B. Reading and the awareness of linguistic segments. *Journal of Experimental Child Psychology*, 1974, **18**, 201–212.

MASSARO, D. Perceptual units in speech recognition. *Journal of Experimental Psychology*, 1974, **102**(2), 199–208.

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., SEGUI, J., & FRAUENFELDER, U. The role of the syllable in language acquisition and perception. In T. F. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representations of speech*. Advances in Psychology series. Amsterdam/New York: North-Holland, 1980.

MERMELSTEIN, P. Automatic segmentation of speech into syllabic units. *Journal of Acoustical Society of America*, 1975, **58**(4), 880–883.

MILLS, C. B. Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior*, 1980a, **19**, 75–83.

MILLS, C. B. Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, **6**(3), 528–535.

MORAIS, J., CARY, L., ALEGRIA, J., & BERTELSON, P. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 1979, **7**, 323–331.

RUBIN, P., TURVEY, M. T., & VAN GELDER, P. Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception and Psychophysics*, 1976, **19**(5), 394–398.

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

SWINNEY, D. A., & PRATHER, P. Phonemic identification in a phoneme monitoring experiment: The

variable role of uncertainty about vowel contexts. *Perception and Psychophysics,* 1980, **27**(2), 104–110.

VAISSIÈRE, J. Speech recognition as models of speech perception. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech. Advances in Psychology series.* Amsterdam/New York: North-Holland, 1980.

## REFERENCE NOTES

1. SEGUI, J., FRAUENFELDER, U., & MEHLER, J. Phoneme monitoring, syllable monitoring and lexical access, in preparation.
2. MARSLEN-WILSON, W. D. *Sequential decision process during word recognition.* Paper presented at the Psychonomic Society meetings in San Antonio, Texas, November 1978.

(Received August 27, 1980)