ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

SPEECH SEGMENTATION IN DUTCH: NO ROLE FOR THE SYLLABLE

[1]*Jean Vroomen and* [2]*Beatrice de Gelder*

[1]Tilburg University, The Netherlands
[2]Tilburg University, The Netherlands and Université Libre de Bruxelles, Belgium

ABSTRACT

Three monitoring experiments are reported investigating the role of the syllable in spoken Dutch. Subjects detected prespecified targets (e.g., *daa* or *daal*) that did or did not correspond to the initial syllable of a word (*daa.ling* and *daal.der*) or pseudoword (*daa.les* and *daal.sel*). In addition, carrier words varied in whether they had clear or ambisyllabic structure (e.g., *ko[rr]el* versus *kor.ting*). No evidence for a syllabic segmentation routine was found. These findings are consistent with the absence of a syllable effect in English which is, like Dutch, another stress-based language.

I. INTRODUCTION

Speech recognition entails the mapping of a continuous time-varying acoustic signal onto discrete linguistic units that mediate between the acoustic signal and the lexicon. Among possible candidates for intermediate units, the syllable has received empirical support, at least in some languages.

In an initial study by [1] it was observed that subjects responded faster to a target if it corresponded exactly to the initial syllable of a target-bearing carrier word. For instance, the target *pa* was detected faster in *pa.lace* than in *pal.mier*, whereas *pal* was detected faster in *pal.mier* than in *pa.lace*. The results were interpreted as showing that the syllable constitutes an intermediate unit in speech perception. Similar results were obtained with other Romance languages like Catalan [2]. For English, however, the syllabic effect could not be replicated. In a study by [3], CV or CVC targets were used as in the original French experiment, but no trace of syllabification was found as the monitoring times for CV and CVC targets were almost identical in both types of carrier words.

Another important finding was that English subjects did not show a syllabic effect when listening to French, whereas French subjects did show a syllabification effect when listening to English material with clear syllable boundaries (as in *bal.cony*). This complete absence of syllabic effects in English subjects was interpreted as evidence for a language-specific processing strategy. The notion is that English subjects do not use a syllabic segmentation procedure because of the many cases of ambisyllabicity in English. For instance, in a word like *ba[ll]ance*, the /l/ belongs neither clearly to the first, nor to the second syllable. Such an intervocalic consonant preceding a short stressed vowel is treated in English as ambisyllabic as it is considered to be part of the first and the second syllable. The syllabic segmentation routine would presumably fail or be more time consuming in such cases.

The hypothesis that the phonological nature of the language determines processing routines of the listener prompts the question whether similar results can be obtained in other languages with unclear syllable structures. Is it, for instance, always the case that languages with widespread ambisyllabicity do not promote syllabic segmentation strategies? A recent study by [4] seems to contradict this assertion. The authors report evidence for syllabic effects in Dutch, which is, like English, a language with unclear and ambisyllabic syllable boundaries. As in the English experiments, subjects detected CV or CVC targets in clear or ambisyllabic target-bearing carrier words. Although the interaction between target and carrier word - which is usually interpreted as a syllabic effect - was not significant, a general CVC target advantage over CV targets was observed. This result was understood as a syllabic effect, because the additional assumption was made that the first syllable of ambisyllabic words in Dutch include the intervocalic consonant. Thus, it was argued that *effectively* the first syllable of the ambisyllabic CV[C] words had in fact CVC structure. However, a potential problem, as the authors admitted, is that the CV targets did not have their own control. Moreover, most carrier words in the reported experiments were inflected, derived, or compounded so that syllabic effects were possibly confounded with morphological complexity such as affix stripping or overlap in stem between target and carrier. However, there were also trials in which the target-bearing words had long vowels (VV) with clear syllabic structure. In these cases, the traditional interaction between target and carrier word was significant. But, this experiment deviated from the usual procedure in that the CVVC target did not correspond to a *segment* of the CVVC target-bearing word, but to the carrier word itself. For instance, the CVVC target *maag* had to be detected in the target-bearing word *maag* (stomach). So, the task in this condition amounts to word-spotting, and not segment detection. Since the interaction was, at least partly, based on fast reaction times in this condition, one needs to control for this factor.

EXPERIMENT 1

In the first experiment we tried to investigate whether Dutch listeners apply a syllabic segmentation routine using the standard procedure for selecting items [1]. Subjects monitored for CV or CVC targets which always corresponded to *segments* (syllabic or not) of a carrier word. There was no morphemic overlap between targets and carriers.

II. METHOD

Subjects: Twenty-eight subjects participated in the experiment. All were students and native speakers of Dutch who were paid a small amount for their participation.

Materials: Twenty-seven pairs of spoken items were used as critical target-bearing words. The stem of the two words of each pair was different, but they shared the three initial phonemes. Seven word pairs had a long vowel (referred to as VV), and twenty a short one (V). Of the word pairs with long vowels, one member of the pair had CVV structure (e.g., *daa.ling*, landing), the other CVVC (*daal.der*, half-a-crown). Of the word pairs with short vowels, one member was ambisyllabic (e.g., *ko[rr]el*, grain), the other had a clear syllable boundary after the first consonant in the cluster (e.g., *kor.ting*, discount). All target-bearing words were bisyllabic lexical nouns with stress on the first syllable.

Two blocks, A and B, of 27 lists each were made. Lists ranged from four to ten words in length. In each list, the two target-bearing words of a word pair appeared one after the other with one to four filler words in between each of them. The order of the experimental and filler words was the same in the two blocks, but the target was changed from one block to the other. Thus, if a subject had to monitor for KO in KORREL and KORTING in block A, another subject from block B would have to monitor for KOR in KORREL and KORTING. The fillers were, as the targets, bisyllabic nouns bearing stress on the first syllable. None of the fillers had the same initial consonant or vowel as the target (i.e. no catch trials).

Procedure: All items were recorded by a native male speaker of Dutch. The words were digitised at a sampling frequency at 20 kHz. Inaudible timing pulses were set under visual control at the onsets of the experimental words. Separate digital audio-tapes were made for the A and B versions of the experiment. Each trial started with a 200 msec warning tone. Then the target was heard twice. The targets were also written down on a prepared sheet so that the subject could read the target. After 3 sec, the first word of a list was presented, followed by the next word after 500 msec. The timing pulses placed at the critical words triggered a timing module that registered subjects' push button response. The next trial started after 8 sec.

Each subject was tested individually in a sound proof room. The items were played back via a Philips 850 DAT recorder over loudspeakers at a comfortable listening level. Subjects were instructed to monitor for the occurrence of the prespecified target in the spoken words. They were told that each list contained two critical items to which they had to respond. Subjects pressed a single button in front of them to indicate a target detection. Each test session lasted about ten minutes.

III. RESULTS and DISCUSSION

Mean reaction times for each subjects and each item were computed. Reaction times longer than 1000 ms were discarded and then cutoff points were established, for each subject and item, at plus or minus two standard deviations. Overall analyses of variance were performed with Target and Carrier as factors on the means for subjects and items, separately for words with long and short vowels. The mean reaction times (from the subject analyses) for each condition are shown in table 1 and 2.

TABLE 1
Mean RTs (msec) for Short Vowels

	CV[C] carrier *ko[rr]el*	CVC carrier *kor.ting*
CV target *ko*	438	439
CVC target *kor*	427	425

For carriers with short vowels, the ANOVAs showed that there was a significant effect of target type in the subjects analysis [$F_1(1,27) = 7.64$, p =

.01], indicating that CVC targets were responded 13 msec faster than CV targets. This effects was, however, not significant in the item analysis [$F_2(1,38) = 1.08$, $p = .30$]. No other effects were significant (all $p > .10$).

TABLE 2
Mean RTs (msec) for Long Vowels

	CVV carrier *daa.ling*	CVVC carrier *daal.der*
CVV target *daa*	486	479
CVVC target *daal*	483	467

For words with long vowels, there is an overall advantage of 12 msec for CVVC carriers over CVV carriers. This effect is marginally significant in the subject analysis [$F_1(1,27) = 2.89$, $p < .10$], but not in the item analysis [$F_2 < 1$]. All other effects were non-significant.

To summarize, with short vowels there was, as in [4], a trend that CVC targets were responded faster than CV targets, but this was not significant in the item analysis. With long vowels, there was no sign of the classical interaction between target and carrier. One possible explanation for this negative outcome is that subjects were, in this particular experiment, induced not to use a syllabic segmentation strategy because of the mixed nature of the item set (most carrier words had an ambisyllabic syllable structure, but a few had clear syllable boundaries). The varied item set might thus prevent subjects from adopting a syllabic segmentation routine. We therefore conducted two other experiments in which syllabic structure was constant. In experiment 2 all items were ambisyllabic, whereas in experiment 3 all items had clear syllable structure. To increase to number of items, pseudowords were used.

EXPERIMENT 2

Experiment 2 was similar to the previous one, except that all items were pseudowords with ambisyllabic structure. If listeners use a syllabic routine, one should find, as in [4], a CVC target advantage because CV[C] words are effectively segmented as CVC.

IV. METHOD

Subjects: Twenty-seven subjects were paid for their participation.

Materials and Design: All details were as in experiment 1, except that the items consisted of 32 pseudoword pairs. One of the members had ambisyllabic CV[C]VC structure (e.g., *ma[ss]er*), the other had clear CVC.CVC structure (e.g., *mas.tel*).

V. RESULTS and DISCUSSION

Mean reaction times are presented in table 3. No traces of syllabic effects were observed since neither the main effect of target type, nor the interaction between target type and carrier approached significance.

TABLE 3
Mean RTs (msec) for Pseudowords with Short Vowels

	CV[C] carrier *ma[ss]er*	CVC carrier *mas.tel*
CV target *ma*	403	404
CVC target *mas*	401	408

ANOVAs with subjects and items as random factors showed that the effects of target, carrier, and the interaction between target and carrier were all non-significant (all $p > .10$).

EXPERIMENT 3

Experiment 3 was similar to the previous one, except that all items were pseudowords with clear syllabic syllable structure. This time, a syllabic segmentation routine should be reflected in a cross-over interaction between target and carrier.

VI. METHOD

Subjects: Twenty-six subjects were paid for their participation.

Materials and Design: As in experiment 2, except that the items consisted of 30 pseudoword pairs. One of the members had a syllabic CVV.CVC structure (e.g., *lee.ten*), the other had clear CVVC.CVC structure (e.g., *leet.sen*).

VII. RESULTS and DISCUSSION

Mean reaction times are presented in table 4. As can be seen, again no syllabic effect was observed.

TABLE 4
Mean RTs (msec) for Pseudowords with Long Vowels

	CVV carrier *lee.ten*	CVVC carrier *leet.sen*
CVV target *lee*	378	372
CVVC target *leet*	381	375

Responses were somewhat faster (3 msec) to CVV targets than to CVVC targets. This was significant by items [$F_2(1,58) = 4.40$, $p < .05$], but not by subjects. All other effects were non-significant.

VIII. CONCLUSION

Three monitoring experiments investigated whether the syllable plays a role in speech segmentation of Dutch. When there is no possible confounding with morphological complexity - like inflections, compounds or overlap in stem between target and carrier - no evidence was obtained for the idea that Dutch listeners are sensitive to syllabic structure. In unambiguous as well as ambisyllabic words or pseudowords, listeners did not show any sign of sensitivity to syllabic structure. These results are in line with the proposal by [3] who argued that listeners of languages in which syllable structure is unclear do not use the syllable as segmentation unit. Instead it is proposed that the metrical distinction between strong and weak syllables is relevant [5]. This suggestion has recently been confirmed for Dutch [6,7]. It thus seems that segmentation routines are critically dependent on the listeners native language in the sense that language-specific phonological information governs processing routines for structuring the speech input of the listener.

ACKNOWLEDGEMENTS

The research was partly supported by a grant from the Human Frontier of Science Programme "Processing consequences of contrasting language phonologies" and by the Ministry of Scientific research of the French-speaking Community of Belgium (Concerted Research Action 91/96 - 148). The research of the first author has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences.

REFERENCES

[1]: Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, **20***, 298-305.

[2]: Sebastian-Galles, N., Dupoux, E., & Mehler, J. (1991). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, **31***, 18-32.

[3]: Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, **25***, 385-400.

[4]: Zwitserlood, P., Schriefers, H., Lahiri, A., & van Donselaar, W., (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory, and Cognition, **19***, 260-271.

[5]: Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, **14***, 113-121.

[6]: Vroomen, J., & de Gelder, B. (in press). Metrical segmentation and lexical competition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance.*

[7]: Vroomen, J. & de Gelder, B. (submitted). Cues to speech segmentation: evidence from juncture misperceptions and word spotting.