A Note on the Role of Phonological Expectations in Speech Segmentation

ANNE CUTLER, DENNIS NORRIS, AND JOHN N. WILLIAMS

MRC Applied Psychology Unit, Cambridge, United Kingdom

Word-initial CVC syllables are detected faster in words beginning consonant–vowel–consonant–vowel (CVCV-) than in words beginning consonant–vowel–consonant–consonant (CVCC-). This effect was reported independently by M. Taft and G. Hambly (1985, Journal of Memory and Language, 24, 320–335) and by A. Cutler, J. Mehler, D. Norris, and J. Segui (1986, Journal of Memory and Language, 25, 385–400). Taft and Hambly explained the effect in terms of lexical factors. This explanation cannot account for Cutler et al.’s results, in which the effect also appeared with nonwords and foreign words. Cutler et al. suggested that CVCV- sequences might simply be easier to perceive than CVCC- sequences. The present study confirms this suggestion, and explains it as a reflection of listener expectations constructed on the basis of distributional characteristics of the language. © 1987 Academic Press, Inc.

Listeners detect a word-initial syllable such as tal faster in words like tally or talcon (which begin consonant–vowel–consonant–vowel) than in words like talcum (which begin consonant–vowel–consonant–consonant).

This finding has been independently reported in this journal by Taft and Hambly (1985) and by Cutler, Mehler, Norris, and Segui (1986). The effect is robust. Taft and Hambly found it in two separate experiments, Cutler et al. in three. All experiments used English-speaking listeners.

Two radically different explanations were offered for the effect. Taft and Hambly ascribed it to the influence of orthographic structure on phonological representations in the lexicon. Taft (1979) proposed the concept of basic orthographic syllable structure (BOSS) whereby special lexical status is accorded to a word’s orthographically defined initial syllable; this consists of the onset, the first vowel, and as many of the consonants following this vowel as could legally constitute a syllabic coda. Thus the BOSS of the word tantrum is tant; the third medial consonant does not belong to the BOSS because tantr is not a legal English syllable. Similarly, the BOSS of tally and talon is tal, while the BOSS of talcum is talc. Taft and Hambly argued that the target tal is detected faster in tally than in talcum because in tally the target corresponds exactly to the word’s BOSS, whereas in talcum the target constitutes less than the BOSS. Therefore, they argued, orthographic structure, in the form of special status for the BOSS, plays a role in the lexical representations accessed from auditory input as well as from visual input.

The significant characteristic of Taft and Hambly’s (1985) account is that it invokes the characteristics of lexical representations, and hence can only be applied to responses which are made after the word has been recognized.

Cutler et al. (1986), however, explained
the faster detection of *tal* in *talon* than in *talcum* quite differently, and without invoking lexical characteristics. They suggested that some phonological patterns might simply be easier to process than others. In particular, they pointed out, alternating sequences of vowels and consonants might be particularly easy to recognize, since there is independent evidence that vowels are easier to identify if they are bounded by consonants (Strange, Verbrugge, Shankweiler & Edman, 1976), while consonants are easier to identify if they are bounded by vowels (Liberman, Delattre, Cooper & Gerstman, 1954). Therefore, words beginning CVCC-, like *talan*, would simply be intrinsically easier to perceive than words beginning CVCC-, like *talcum*.

This explanation can be applied to syllable detection responses which occur prior to lexical retrieval. A prelexical account of the effect was necessary to cope with Cutler et al.’s (1986) finding that monolingual English listeners showed a response time advantage for CVCC- over CVCC- structures even when the stimuli were French words or nonsense words, neither of which could activate lexical representations. Cutler et al.’s subjects must have been responding prior to lexical access, in at least these two of their three experiments.

Similarly, though, Taft and Hambly’s (1985) subjects were demonstrably responding after lexical access, since response times in their experiment were faster to high frequency than to low frequency words. Taft and Hambly did not offer an explanation of why subjects in their experiment responded postlexically while subjects in other syllable detection experiments respond prelexically (Taft & Hambly cite Mehler, Dommergues, Frauenfelder & Segui, 1981, and Segui, Frauenfelder & Mehler, 1981, in this context). They established that the difference could not be due to the fact that in their experiment subjects were instructed to listen for syllables occurring anywhere in the word rather than, as in most other experiments, only in word-initial position; when they varied their instructions (in a study which their paper does not report in detail) to confine responses to word-initial syllables only, the frequency effect remained, indicating that subjects were still responding postlexically. The phonological structure effect, i.e., the faster detection of *tall* in *tally* than in *talcum*, also replicated. There is one clear difference between Taft and Hambly’s methodology and that of most other syllable detection experiments, however, which does account for the difference in response level: Taft and Hambly required yes–no responses to every item. Most syllable detection studies (including Mehler et al.’s, Segui et al.’s and Cutler et al.’s, 1986) require a go/no-go response, which encourages faster responding. Taft and Hambly’s response times were indeed some 300 ms slower on average than those in the Cutler et al. experiments, and it is reasonable to suppose that the extra processing involved by their methodology delayed responses to a point where lexical effects could become apparent.

Postlexical responding, however, does not necessarily imply a lexical source for the phonological structure effect. A simple appeal to parsimony suggests that having two explanations for the same effect is undesirable; if one of the two could explain all occurrences of the effect, it might be considered preferable to the other. Taft and Hambly’s (1985) lexically based account clearly cannot explain Cutler et al.’s (1986) finding of faster responses to CVCC- than to CVCC- structures when the stimuli were foreign words or nonsense. Cutler et al.’s ease of processing hypothesis could, however, explain Taft and Hambly’s result. The phonological structure of a word must be processed before lexical access can take place, and there is no reason why facilitation at this early stage should not carry
through and result in faster responses even if those responses are (for whatever reason) delayed until after lexical access has taken place.

Therefore there is a clear reason to prefer Cutler et al.'s (1986) explanation of the phonological structure effect to Taft and Hambly's (1985), on grounds of simplicity alone.

Unfortunately some recent results have cast doubt upon the generality of Cutler et al.'s (1986) explanation. Cutler, Butterfield, and Williams (1987) investigated a frequently reported finding that detection of word-initial phonemes (e.g., [b]) is faster if the phoneme precedes a vowel (as in *boom*) than if it is part of a cluster (as in *broom* or *bloom*). Cutler, Butterfield, and Williams suggested that this finding might also reflect the intrinsically easier segmentability of alternating sequences of consonants and vowels. However, they discovered that detection of stop consonants in consonant clusters could actually occur faster than detection of the same consonants in CV sequences if listeners were led to expect clusters.

Therefore it is not the case that alternating sequences of consonants and vowels are necessarily always easier to perceive than sequences in which consonants occur together. But the results of the Cutler, Butterfield, and Williams (1987) study suggest an alternative basis for the Cutler, Mehler, Norris, and Segui (1986) ease of processing hypothesis. Cutler, Butterfield, and Williams showed that listeners who expected clusters detected phonemes faster in clusters, while listeners who expected CV sequences detected phonemes faster in CV sequences. If listeners in syllable detection experiments are similarly forming expectations as to the phonological structure of the stimuli, then they should find stimuli which conform to those expectations rather easier to process than stimuli which violate their expectations. Thus a modified version of the Cutler, Mehler, Norris, and Segui proposal would suggest that words beginning CVCV- are easier to process than words beginning CVCC- because words beginning CVCV- are what listeners are expecting to hear.

In fairness, though, a modified version of the Taft and Hambly (1985) hypothesis could also be constructed which could offer a more complete account of the available data. As formulated by Taft and Hambly, the BOSS hypothesis is necessarily post-lexical and hence cannot account for Cutler, Mehler, Norris, and Segui's (1986) phonological structure effect. But suppose that syllable detection is facilitated when the syllable target corresponds to the maximal possible initial syllable allowed for that word by the phonological laws of English. There is no necessity to assume lexical involvement in this effect, let alone orthographic influence at this early phonological processing stage.

We might call this the hypothesis of maximal allowable syllable structure (MASS). The MASS of *tally* and *talon* is *tal*, whereas the MASS of *talcum* is *talc*, which could be, and indeed in this instance is, a syllable of the language by itself.

This hypothesis can be tested and distinguished from the ease of processing hypothesis. Consider two words beginning CVCC-: In one the two adjacent consonants can together constitute a syllable coda, making the MASS CVCC- *Talcum* is such a word. In the other word, however, the two adjacent consonants cannot combine in a syllabic coda. The British place name *Talgarth* is such a word: there are no English syllables ending with the segments [lg], so the MASS of *Talgarth* is *tal*. The MASS hypothesis predicts that *tal* should be detected faster in *Talgarth* than in *talcum*. The ease of processing hypothesis, however, predicts that since both *talcum* and *Talgarth* begin CVCC-, they should be equally easy (or difficult) to process.

The difference predicted by the MASS hypothesis would of course also be predicted by Taft and Hambly's (1985) BOSS hypothesis. However, Taft and Hambly
were not in a position to test it. This is be- dure. In order to control for intrinsic diffi-
cause the English language does not pro­
vide materials for the carefully matched  
pairs which Taft and Hambly’s method­
ology required. There are enormous  
numbers of English words beginning  
CVCC-, in which the adjacent consonants  
do not allow a syllabic coda interpretation (nutmeg, magpie, picnic, bodkin, cutler,  
etc.), just as there are enormous numbers  
of words beginning CVCC- where the two  
adjacent consonants can be interpreted as a  
syllabic coda (talcum, hectic, gospel, com­
fort, salvage, etc.). But there are virtually  
no such pairs which share the initial syl­
uble (and which also satisfy the morpho­
logical constraints which Taft and Hambly  
imposed on their materials); pairs can only  
be constructed using proper names (Tal­
garth and talcum), morphologically com­
plex words (midwife and midget), or infre­
quently words (lanyard and lancet).

This problem does not arise, of course,  
for a test of the MASS hypothesis. In order  
for this hypothesis to be applied to the  
Cutler, Mehler, Norris, and Segui (1986)  
findings, it must demonstrably be appli­
cable to prelexical responses. Therefore a  
test of the hypothesis must force subjects  
to respond prelexically, and the best way to  
do that is to present them with nonsense  
words.

Nonsense words can be constructed in  
pairs which satisfy the relevant criteria.  
For instance, take the nonsense word pair  
tafgarp and taftarp. The MASS of tafgarp  
is taf, because there are no English syl­
lables ending [fg]. The MASS of taftarp is  
taft, because taft is a possible English syl­
lable. The MASS hypothesis predicts that  
taf will be detected faster in tafgarp (where  
it corresponds exactly to the item’s MASS)  
than in taftarp (where it constitutes less  
than the MASS).

In our first experiment we tested the  
MASS hypothesis in this way. In order fur­
ther to mimic the conditions of the Cutler,  
Mehler, Norris, and Segui (1986) experi­
ments, we used a go/no-go detection proce-

these groups, half the subjects heard the lists in AB, half in BA order.

Results

The mean response times are presented in Table 1. The results do not support the prediction of the MASS hypothesis; response times to CVC targets like *taf* are not faster in *tafgarp* items than in *taftarp* items. Nor do response times to CV targets like *ta* differ as a function of phonological structure; post hoc analyses showed that responses to *tafgarp* and *taftarp* did not differ significantly for either target type separately. These results are as predicted by the Cutler, Mehler, Norris, and Segui (1986) hypothesis, which holds that *tafgarp* and *taftarp* should be equally difficult to process, irrespective of size of syllable target.

The only significant result from this experiment was that CVC targets like *taf* were detected faster than CV targets like *ta* ($F_1(1,16) = 28.35, p < .001; F_2(1,40) = 78.07, p < .001$). This effect did not interact with the MASS structure variable.

The failure to find an effect of phonological structure with either target type rules out the possibility that the null result in the CVC target condition could be due to simultaneous operation of a MASS effect and some unspecified effect working in the opposite direction (such as a difference in clarity of articulations, or in length, between *tafgarp* and *taftarp* items). Any such effect should show up as an advantage for *taftarp* items in the CV target condition, in which the MASS factor is controlled; but as Table 1 clearly shows, there was no such advantage.

Discussion

This experiment failed to support the MASS hypothesis. There is no evidence that detection time for syllable targets in English is affected by phonological constraints on syllabification.

Instead, we suggest that the previous findings of faster detection times for CVC targets in CVCV- as opposed to CVCC- structures can most simply be explained by a variant of the hypothesis that Cutler, Mehler, Norris, and Segui (1986) proposed. CVCV- structures are easier to process in a syllable detection task than CVCC- structures. They are easier to process because they are what subjects expect, once they have been given the syllable target.

The notion that subjects are forming expectations about the phonological structure of the stimuli is in fact directly supported by the results of Experiment 1. CV targets were harder to detect than CVC targets. In this experiment we used five different vowels (in contrast to the Cutler, Mehler, Norris, and Segui (1986) experiments, in which the vowel was always *a*). It is likely that with visual presentation CV targets encourage the formation of more varying expectations as to phonological structure than do CVC targets. That is, some subjects seeing TA- might be led to expect a word with the short vowel [tæ], while others might expect long vowels such as [ta] or [te]. Experiment 2 tested this hypothesis.

### Table 1

<table>
<thead>
<tr>
<th>Target type</th>
<th>CV (e.g. <em>ta</em>)</th>
<th>CVC (e.g. <em>taf</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonological</td>
<td>CVC MASS (e.g. <em>tafgarp</em>)</td>
<td>419</td>
</tr>
<tr>
<td>Structure</td>
<td>CVCC MASS (e.g. <em>taftarp</em>)</td>
<td>435</td>
</tr>
</tbody>
</table>

Ten members of the Applied Psychology Unit community read out loud the 24 CVC targets and 17 CV targets from Experiment 1. (There were only 17 CV targets because some CVC targets, e.g. *taf* and *tan*, shared initial CV sequences). Table 2 presents the distribution of vowel qualities produced by these speakers. It can clearly be seen that
TABLE 2
Vowel Quality of Subjects' Productions, Experiment 2 (Proportions)

<table>
<thead>
<tr>
<th>Target type</th>
<th>CV (e.g. ta)</th>
<th>CVC (e.g. taf)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short vowels</td>
<td>.135</td>
<td>.987</td>
</tr>
<tr>
<td>Long vowels</td>
<td>.865</td>
<td>.013</td>
</tr>
</tbody>
</table>

whereas CVC targets were virtually always produced with short vowels, CV targets were produced with more varying vowel qualities, but mostly with long vowels.

All subjects reported (three of them spontaneously) that the CVC targets were far easier to read out than the CV targets because the final consonant phonologically disambiguated the vowel.

Discussion

CVC targets produced short vowel responses, CV targets produced a majority of long vowel responses. All items in Experiment 1 in fact had short vowels. Therefore it is reasonable to assume that CVC targets allowed subjects to form more accurate expectations as to the actual phonological structure of the stimuli, and that these expectations are the source of the advantage for CVC over CV targets in that experiment.

We assume that subjects similarly formed expectations about the phonological structure of the stimuli in the two syllable detection studies which prompted the present research. For instance, subjects presented with tal (either visually, as in Cutler, Mehler, Norris, and Segui's (1986) studies, or auditorily, as in Taft and Humbly's, 1985) might simply be more likely to expect words beginning CVCV-than words beginning CVCC-. We tested this suggestion in the following two experiments.

EXPERIMENT 3

Forty-four members of the Applied Psychology Unit community, none of whom had participated in Experiment 2, provided polysyllabic word completions for the visual targets used by Cutler, Mehler, Norris, and Segui (1986). (It was not possible to use the Experiment 1 targets, since a majority of them had no real word completions). Half the subjects completed BA-, MA-, TA-, CA-, SAL-, PAL-, and GAL-, while the other half completed BAL-, MAL-, TAL-, CAL-, SA-, PA-, and GA-. The items were embedded in a list of 20 items in all. Table 3a shows that both types of target produced vastly more completions with CVCV- structure than with CVCC- structure. Table 3b further confirms the findings of Experiment 2: CVC targets produced a very great preponderance of completions with short vowels over completions with long vowels (more than 20:1) whereas CV targets produced much more varied phonology (the ratio of short vowels to long being less than 2:1).

EXPERIMENT 4

Fifty-six members of an introductory linguistics class provided polysyllabic word completions for the auditory targets used by Taft and Hambly (1985). As in Taft and Hambly's experiments, the targets were spoken in isolation, and each was presented twice in succession.

TABLE 3
(a) Phonological structure of word completions, Experiment 3 (proportions)

<table>
<thead>
<tr>
<th>Target type</th>
<th>CV (e.g. ta)</th>
<th>CVC (e.g. tal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVCV-</td>
<td>.799</td>
<td>.779</td>
</tr>
<tr>
<td>CVCC-</td>
<td>.201</td>
<td>.221</td>
</tr>
</tbody>
</table>

(b) Vowel quality in word completions, Experiment 3 (proportions)

<table>
<thead>
<tr>
<th>Target type</th>
<th>CV</th>
<th>CVC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short vowels</td>
<td>.656</td>
<td>.961</td>
</tr>
<tr>
<td>Long vowels</td>
<td>.344</td>
<td>.039</td>
</tr>
</tbody>
</table>
Table 4 shows that CVCV- completions predominated once again. Forty subjects produced a majority of CVCV- completions, and only ten a majority of CVCC-completions (the remaining six producing equal numbers of each). This difference was highly significant ($t(55) = 4.76, p < .001$).

Of the 20 targets, 13 were given a majority of CVCV- completions. The phonological structure effect was not significant over all 20 items ($t(19) = .92, p > .3$). Of the remaining 7 targets, 4 were notable for receiving a very large majority of CVCC-completions. These were vin, sim, hec, and ven. (Without these four items, the phonological structure effect is significant: $t(15) = 3.14, p < .01$.) It is an impressive fact that all of these four targets also failed to produce the phonological structure effect for Taft and Hambly (1985, p. 333).

Discussion

Experiments 3 and 4 clearly showed that subjects think of CVCV- words rather than CVCC- words when they are presented with syllable targets.

It is reasonable to assume, therefore, that the response time advantage for words like talon over words like talcum in Cutler, Mehler, Norris, and Segui's (1986) experiments arose because the talon words conformed more precisely to subjects' phonological expectations about the stimuli than the talcum words did.

Likewise, it is reasonable to assume that the subjects in Taft and Hambly's (1985) experiment constructed the same sort of phonological expectations about the stimuli, and that they responded to tally faster than talcum because tally more closely corresponded to their expectations; the facilitatory effect at the phonological processing level carried through even though Taft and Hambly's subjects did not respond until after they had recognized the words.

Why should subjects be more likely to expect words of CVCV- than CVCC- structure? The answer to that question probably lies in simple frequency of occurrence. First, English contains more words of the CVCV- type. David Carter, of the University of Cambridge Computer Laboratory, searched a computerized dictionary of 38,400 English words on our behalf. There were 11,026 polysyllables beginning CVCV-, but only 8330 polysyllables beginning CVCC-. But more interestingly, frequency of occurrence is higher for CVCV-polysyllables. A subset of this computerized dictionary, containing the 13,800 most frequent words, lists individual word frequencies (after Kucera & Francis, 1967). In this subset there are 4601 polysyllables beginning CVCV-; they have a mean frequency of 21.8. There are only 2905 polysyllables beginning CVCC-, and they have a mean frequency of 15.9. These figures suggest that on average listeners should hear polysyllables beginning CVCV- more than twice as often as they hear polysyllables beginning CVCC-. The expectation that this ratio should be preserved in an experimental situation is surely a very rational one.

The present research was prompted by the existence of two very different explanations for the same effect, that syllable targets are detected faster in CVCV- versus CVCC- environments. As found by Cutler, Mehler, Norris, and Segui (1986), the effect was serendipitous, and unrelated to the main issues of their study. Cutler, Mehler, Norris, and Segui (1986) suggested that the effect might arise because CVCV- sequences could be easier to perceive, at a prelexical level, than CVCC- sequences. The present research indicates that this explanation is correct, and that the increased perceptibility is due simply to the nature of

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>PHONOLOGICAL STRUCTURE OF WORD COMPLETIONS, EXPERIMENT 4 (PROPORTIONS)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All targets</td>
</tr>
<tr>
<td>CVCV-</td>
<td>.561</td>
</tr>
<tr>
<td>CVCC-</td>
<td>.439</td>
</tr>
</tbody>
</table>
phonological expectations which subjects construct on the basis of their experience with the language.

Taft and Hambly (1985), however, had predicted the phonological structure effect. Their explanation invoked characteristics of the lexical representations of their stimulus words. This postlexical explanation must be rejected, since it cannot account for the fact that Cutler, Mehler, Norris, Segui (1986) found the phonological structure effect with nonwords and foreign words. But the general structure of their explanation, which was based on allowable syllabifications in English, must also be rejected, on the basis of Experiment 1 of the present study. If syllabification mattered at the relevant stage of processing, then differently syllabifiable CVCC- sequences should produce just the same target detection differences as CVCV- versus CVCC-sequences. Experiment 1 showed that this is not the case. Experiments 2 to 4 showed that the phonological structure effect is due to listeners’ expectations: any CVCC- sequence is less expected than a CVCV- sequence.

The lesson is clear. In target detection experiments listeners use their knowledge of the language to construct expectations about the phonological structure of the stimuli. These expectations must be taken into account in the prediction and explanation of experimental effects.

APPENDIX: Stimuli Used in Experiment 1

<table>
<thead>
<tr>
<th>CVC MASS</th>
<th>CVCC MASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>tafgarp</td>
<td>taftarp</td>
</tr>
<tr>
<td>rebfat</td>
<td>rebzat</td>
</tr>
<tr>
<td>biknard</td>
<td>biksard</td>
</tr>
<tr>
<td>lagpoon</td>
<td>lagzoon</td>
</tr>
<tr>
<td>kenlart</td>
<td>kentart</td>
</tr>
<tr>
<td>pamdole</td>
<td>pamphole</td>
</tr>
<tr>
<td>tanlin</td>
<td>tandin</td>
</tr>
<tr>
<td>lonmude</td>
<td>londude</td>
</tr>
<tr>
<td>dinwar</td>
<td>dintar</td>
</tr>
<tr>
<td>bonhaw</td>
<td>bontaw</td>
</tr>
</tbody>
</table>

APPENDIX—Continued

<table>
<thead>
<tr>
<th>CVC MASS</th>
<th>CVCC MASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nanlap</td>
<td>nandap</td>
</tr>
<tr>
<td>dimlup</td>
<td>dimpup</td>
</tr>
<tr>
<td>taplin</td>
<td>tapsin</td>
</tr>
<tr>
<td>dasbale</td>
<td>daskale</td>
</tr>
<tr>
<td>dunlin</td>
<td>duntin</td>
</tr>
<tr>
<td>bamgarl</td>
<td>bamparl</td>
</tr>
<tr>
<td>difgun</td>
<td>difun</td>
</tr>
<tr>
<td>jonlup</td>
<td>jondup</td>
</tr>
<tr>
<td>danled</td>
<td>danded</td>
</tr>
<tr>
<td>ruslole</td>
<td>ruskole</td>
</tr>
<tr>
<td>dufnab</td>
<td>duftab</td>
</tr>
<tr>
<td>kumjite</td>
<td>kumpite</td>
</tr>
<tr>
<td>kaspole</td>
<td>kasdole</td>
</tr>
<tr>
<td>rudbin</td>
<td>rudzin</td>
</tr>
</tbody>
</table>

REFERENCES


(Received October 21, 1986)