THE PERCEPTUAL INTEGRITY OF INITIAL CONSONANT CLUSTERS

Anne Cutler and Sally Butterfield

MRC Applied Psychology Unit, Cambridge

The most inportant component of a syllable is the vowel, because it is the only component that cannot be dispensed with. In some languages, syllables are restricted to very simple structures - only consonant-vowel (CV), for instance. In English, however, syllables can have a wide range of structures. The words <u>are</u> and <u>screeched</u> are both monosyllabic. In prevocalic position, i.e. in the <u>onset</u> of the syllable, English words can have zero, one, two or three phonemes. That is, <u>oak</u>, <u>soak</u>, <u>stoke</u> and <u>stroke</u> are all legitimate English syllables.

In some perceptual tasks, consonants in clusters appear to be treated differently from consonants in isolated prevocalic position. For instance, listeners asked to detect a word-initial target phoneme respond faster if the phoneme precedes a vowel than if it precedes another consonant in a cluster (Foss & Gernsbacher, 1983; Treiman, Salasoo, Slowiaczek & Pisoni, 1982). That is, /b/ takes longer to detect in <u>brand</u> or <u>bland</u> than in <u>band</u>, while /s/ takes longer to detect in <u>stoke</u> or <u>stroke</u> than in <u>soak</u>.

Treiman et al. (1982) argued that this finding reflected listeners' perception of an initial consonant cluster as a coherent unit, the syllabic onset, so that lengthened detection times for phonemes in clusters were caused by the need to segment the unitary cluster into its constituent phonemes. However, it could also reflect the intrinsic ease of segmentation of phonemes as a function of context. Cutler, Mehler, Norris and Segui (1986) found that detection times for word-initial syllables were faster in words beginning CVCV (e.g. <u>balance)</u> than in words beginning CVCC (e.g. <u>balcony)</u>, and suggested that alternating sequences of consonants and vowels might be intrinsically easier to segment into their constitutent phonemes.

In phoneme detection experiments, the phoneme targets are usually specified in a neutral form followed by a word <u>model</u>. For instance, subjects are told to listen for "[ba] as in boy". The word models customarily have single-phoneme onsets. If Treiman et al. are right, and syllabic onsets are perceived as integral units, then perhaps the v*Drd nodels with single-phoneme onsets are leading listeners to expect single-phoneme onsets only. If this is the case, one would predict that sinply providing cluster-onset models would remove the response time advantage for phonemes in single-phoneme onsets. That is, targets on words like <u>band</u> would be responded to faster when the /b/ target is modelled with a single-phoneme onset than when it is modelled with any cluster onset; but targets on words like <u>bland</u> would be responded to faster when the /b/ target is modelled with a /bl/ onset rather-than either a /b/ or a /br/ onset; and targets on words like <u>brand</u> would be responded to faster when the /b/ target is modelled with a /br/ onset rather than either a /b/ or a /bl/

onset. Figure la shows the result predicted by this hypothesis. This result would suggest that in previous experiments, the single phoneme advantage may have been due to the models that were used. Nevertheless, the existence of a matching effect would in itself be consistent with the hypothesis that syllabic onsets can function as integral units. However it is not the result that would be predicted by the hypothesis that consonant-vowel sequences are intrinsically easier to segment than consonant clusters. This latter hypothesis holds that

FIGURE 1
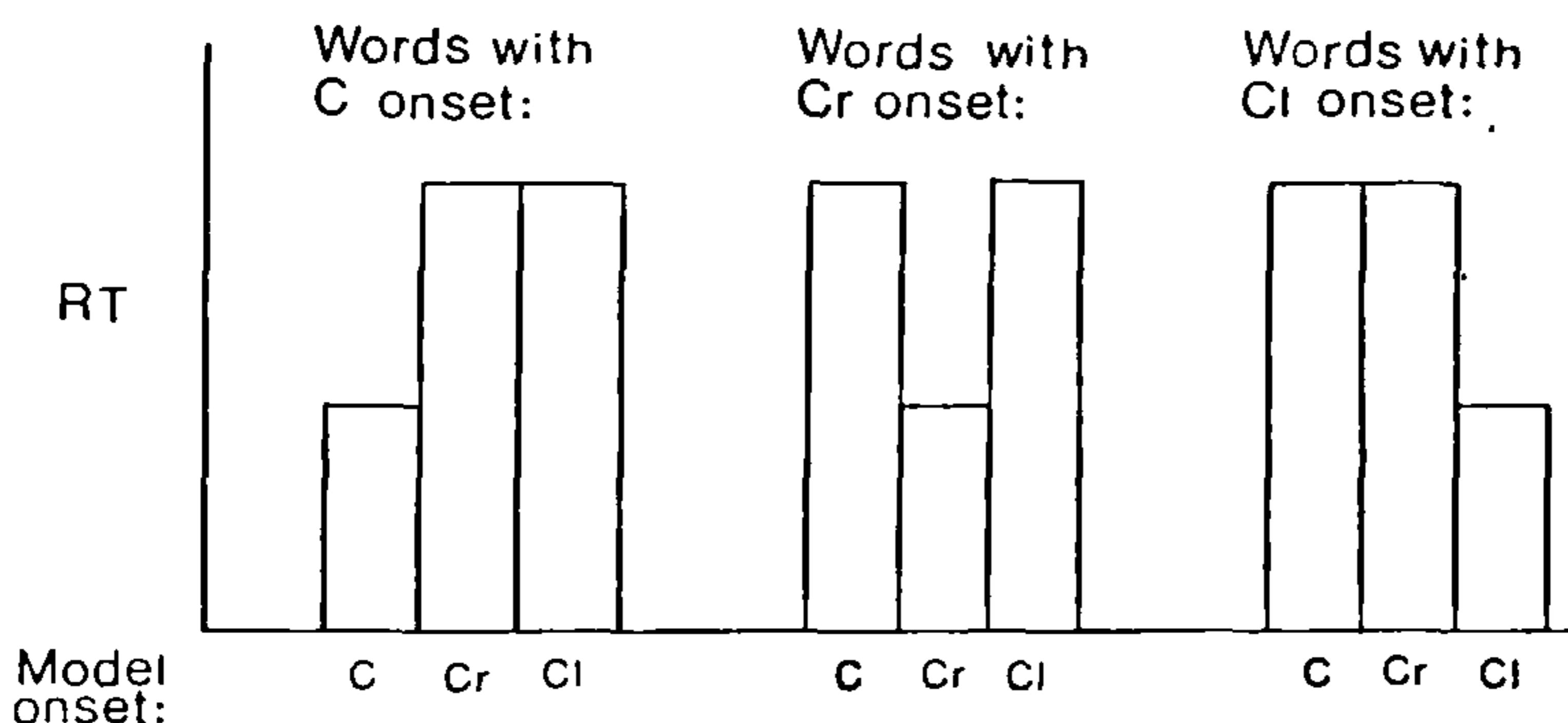
(a) Relative response times predicted by the integral onset expectation hypothesis
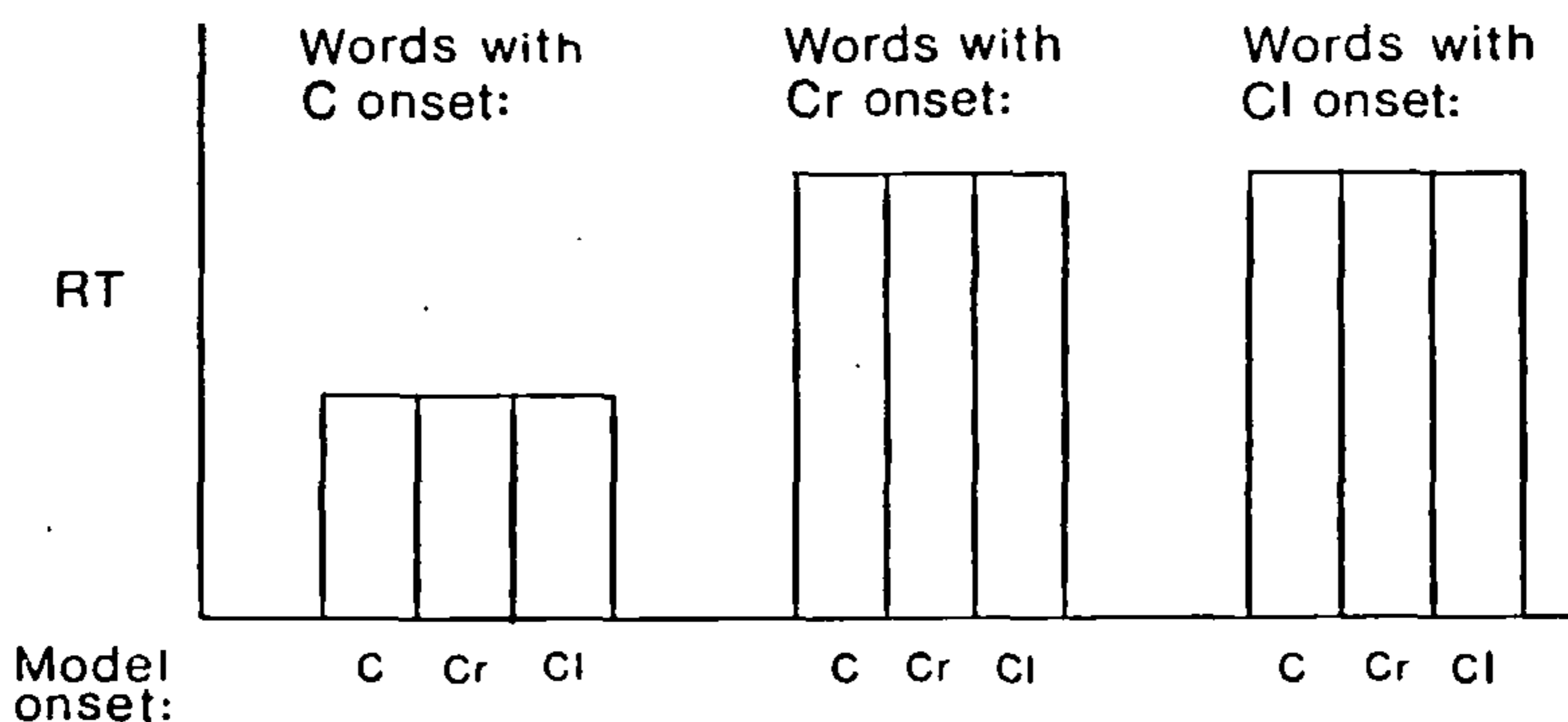
(b) Relative response times predicted by the intrinsic segmentability hypothesis
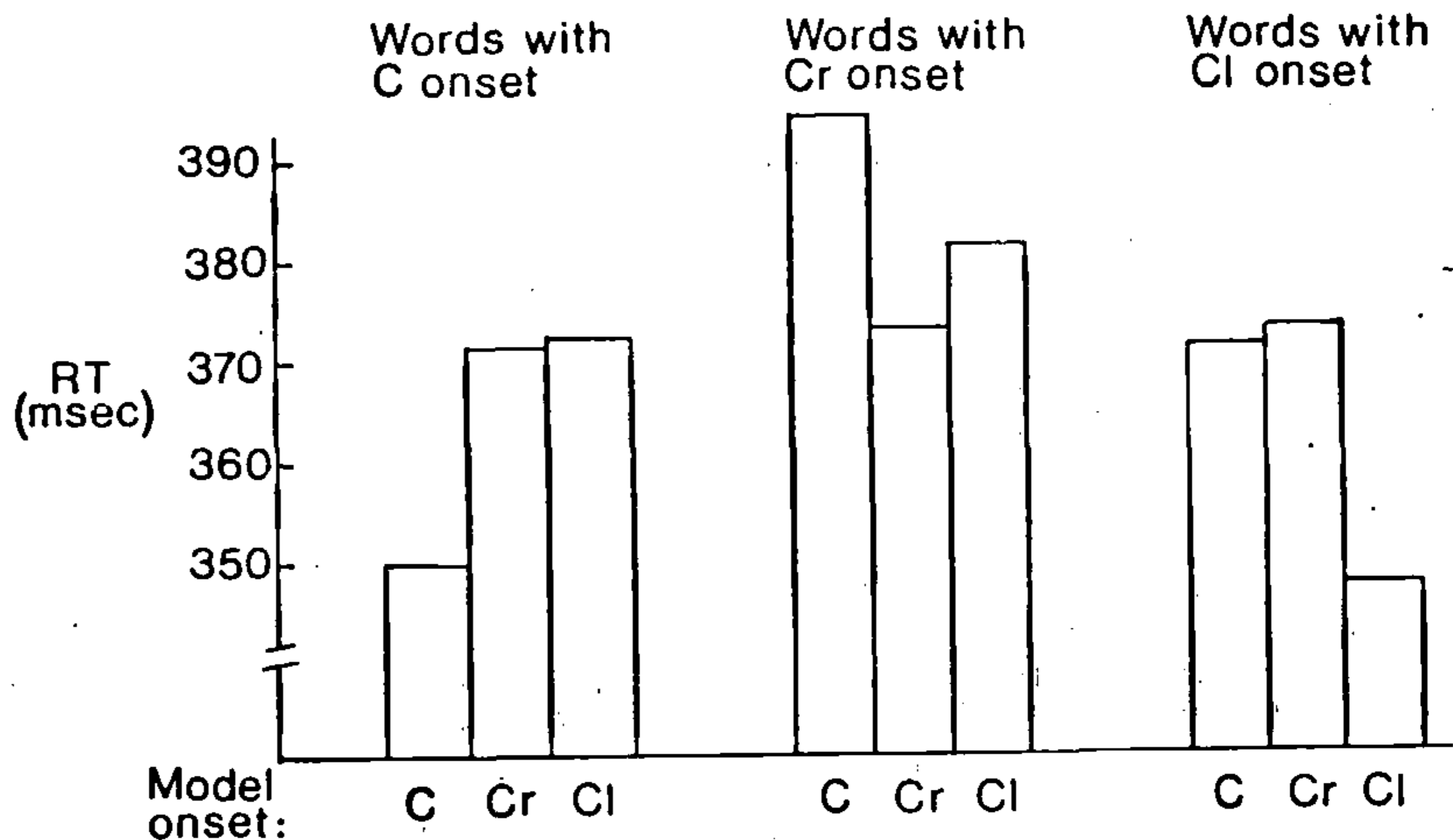
THE PERCEPTUAL INTEGRITY OF INITIAL CONSONANT CLUSTERS

phoneme detection advantages with single-phoneme onsets are due to purely acoustic factors; these are constant across conditions, and should be unaffected by the model which is provided for the phoneme target. This hypothesis clearly predicts a replication of the single-phoneme advantage in the present experiment - responses to targets on words like band should always be faster than responses to targets on words like bland or brand. Figure 1b shows the result predicted by the intrinsic segmentability hypothesis.

In our first experiment, therefore, we presented listeners with standard phoneme-monitoring materials, in which some of the target-bearing words began with single-phoneme onsets while others began with clusters. The design kept variation across conditions to a bare minimum. All listeners heard exactly the same set of sentence materials. All listeners performed the same phoneme detection task on each sentence. The only difference between groups of subjects was in the models which were presented for the various phoneme targets. Each group had one of the three phoneme targets modelled with a single-phoneme onset and the other two targets modelled with cluster onsets.

FIGURE 2

Results of Experiment 1

THE PERCEPTUAL INTEGRITY OF INITIAL CONSONANT CLUSTERS

The target phonemes were /p/, /b/ and /g/. One subject group heard "/p/ as in pink", "/b/ as in blue" and "/g/ as in green"; a second heard "/p/ as in plate", "/b/ as in bowl", and "/g/ as in glass"; a third heard "/p/ as in proud", "/b/ as in brave" and "/g/ as in good". There were eighteen words beginning with each target phoneme, of which six began with a single-phoneme (C) onset, six with a Cr onset, and six with a Cl onset. The words were embedded in sentence contexts.

One recording of the sentences was made, the materials were digitised, and three tapes constructed; they differed only with respect to which set of target models were used. Each sentence was preceded by the appropriate target specification with model. Twenty-one subjects heard each tape. Their response times were collected from target phoneme onset.

Figure 2 shows the mean response times. It can be clearly seen that for each target word onset, the fastest response times are achieved when the model has the same onset. In other words, the pattern of results resembles the prediction in Fig. 1a and differs from the prediction in Fig. 1b. The interaction of word onset and model onset was statistically significant (p<.001).

The results are therefore in accord with the hypothesis that the model itself can set up expectations about syllabic onset. The alternative hypothesis, namely that phoneme targets are necessarily always harder to perceive in clusters than before vowels, because clusters are intrinsically harder to segment than consonant-vowel sequences, can therefore be rejected.

The "matching model" effect is certainly consistent with the hypothesis that initial consonant clusters are perceived as integral units; but it is not direct evidence in its favour. Although the models appear to have influencd subjects' expectations, the expectations may not have been specifically about syllable structure. Subjects may simply have expected certain sequences of phonemes, in which case the effect should be as easy to invoke with, say, onset-plus-nucleus combinations. If the matching model effect is not specific to syllabic onsets, it cannot serve even indirectly as evidence for the integrity of onsets.

In our second experiment, therefore, we attempted to replicate the matching model effect with CV and CVC sequences. As in the preceding experiment, each phoneme target had alternative models; but the models in this case each had the same onset. All the rest of the model word differed. For instance, the two models for /b/ were "/b/ as in Ben" and "/b/ as in Billy". The response items varied in the degree to which they overlapped with the model - in the initial (target) phoneme only, in the initial consonant and its following vowel, or in the whole initial (CVC) syllable.

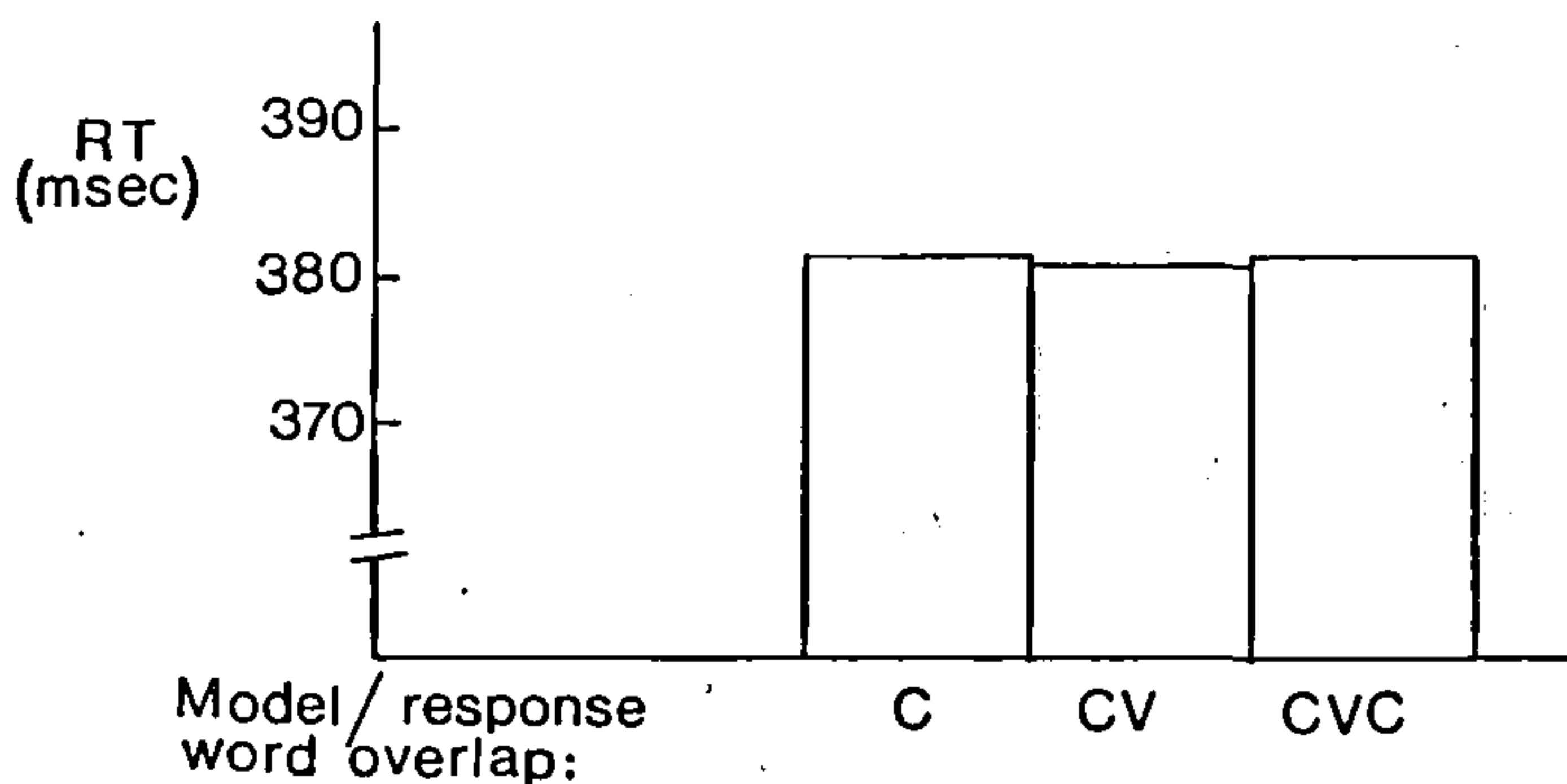THE PERCEPTUAL INTEGRITY OF INITIAL CONSONANT CLUSTERS

There were four target phonemes in this experiment: /p/, /b/, /g/ and /t/. Two target models were chosen for each phoneme; these were all proper names: Pam and Polly, Ben and Billy, Gus and Gary, Tim and Terry. Eight words beginning with each phoneme were chosen. Of these, two began with the same three phonemes as each of the models, while a further two began with the same two phonemes as each of the models. For example, for /t/ the words were timid, timber, terror, terrace, tickle, tissue, tepid and testing. A sentence was constructed for each word. All sentences were recorded once, and as in the preceding experiment, separate tapes were made which differed only in which model was provided for a given phoneme target.

Twenty-five subjects heard each tape, and their detection responses were measured from target onset.

Figure 3 presents the results from this Experiment   When the model matched the target word only on syllabic onset, i.e. on the first phoneme, mean response time was 382 msec; an additional one or two matching phonemes speeded responses by at most one msec.

FIGURE 3

Results of Experiment 2

The matching model effect therefore does not generalise to any phonemic sequence, but appears to be specific to syllabic onsets. We would argue that the matching model effect operates at a fairly abstract level of representation, which is also the level of representation at which the phoneme detection task is performed. Consider that successful detection of a sound such *as* /b/ in words as diverse as barn, beat, brevity, bludgeon, and balloon requires a relatively abstract representation of the target, one which will be impervious to variations in closure duration, burst amplitude and formant transitions which may render the acoustic form representing /b/ in one word quite different from the acoustic form of /b/ in another word. Experimenters using the phoneme-monitoring task have long known this, and with equanimity have presented target specifications in a constant form despite considerable variation across response items. Moreover, the target specification typically emphasises the need to abstract away from acoustic form by first specifying the target in a relatively neutral form, i.e. followed by a schwa, and then modelling it on a particular word. Thus subjects are forced to form an abstract representation of the target in order to perform the detection task successfully regardless of context, and they are encouraged to do so by the way the target is specified. What is most interesting about the present result is that at this abstract level of representation the internal structure of syllables, including the form of the onset, is a relevant feature. We suggest that the matching model effect is a kind of priming effect, and what has been primed is the assignment of syllabic structure at this level of representation.

Note that in everyday language processing abstract representations of this kind, including a representation of syllabic structure, may quite frequently be required. For example, we cannot produce or recognise rhymes without appreciating that the essence of classical rhyme is identity of all segments following the onset of the stressed syllable.

FOOTNOTE

REFERENCES

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

D.J. Foss and M.A. Gernsbacher, Cracking the dual code: toward a unitary model of phoneme identification. Journal of Verbal Learning and Verbal Behaviour, Vol. 22, 609-632, (1983).

R. Treiman, A. Salasoo, L.M. Slowiaczek and D.B. Pisoni, Effects of syllable structure on adults' phoneme monitoring performance. Progress Report No. 8, Indiana University, Speech Research Laboratory, (1982).