MODELLING LEXICAL ACCESS FROM CONTINUOUS SPEECH INPUT

Anne Cutler

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
and
MRC Applied Psychology Unit, Cambridge, U.K.

Dennis Norris

MRC Applied Psychology Unit, Cambridge, U.K.

and

James McQueen

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

MODELLING LEXICAL ACCESS
FROM CONTINUOUS SPEECH INPUT

Anne Cutler

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
and
MRC Applied Psychology Unit, Cambridge, U.K.

Dennis Norris

MRC Applied Psychology Unit, Cambridge, U.K.

and

James McQueen

Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands

1. Introduction: The Language-Specificity of Segmentation Procedures

Continuous speech input is the norm; only relatively rarely do we hear isolated spoken words, and even less often do we hear multi-word sequences in which the individual words are separated one from another by, for instance, pauses. In most spoken language, one word follows on from the next with no intervening pause, and indeed, in the majority of cases with no exploitable cue to the presence of a boundary between any word and the word which follows.

Nevertheless, the recognition of speech must involve *segmentation* of utterances into their component words, since only the component words, not the entire utterance, will be represented in the listener's lexical memory. Despite the rarity of clearly marked word boundaries, however, listeners can rapidly and reliably recognise individual words in utterances in their native language.

Research summarised by many other contributors to this Forum deals with procedures for explicit segmentation of continuous speech input. Since other speakers are dealing with this issue in detail, it will not be elaborated upon here. But the important point to note is that these proposed solutions to the segmentation problem are essentially language-specific. The reason for this is that they exploit aspents of the phonological structure of the input language. Languages differ quite fundamentally in phonological structure, and hence any procedure exploiting the phonology of Language A will only work for languages which share A's phonology (or, more precisely : the relevant aspects of A's phonology). If Language B has a different phonological structure, then the associated segmentation procedure will of necessity also be different.

In the earliest investigations within this line of research, studies with French listeners suggested that they segment spoken utterances into syllables (Mehler, Dommergues, Frauenfelder & Segui, 1981 ; Segui, Frauenfelder & Mehler, 1981). Studies with English listeners, on the other hand, produced evidence of the use of a stress-based segmentation procedure (Cutler & Norris, 1988; Cutler & Butterfield, 1992). Since French does not have English-like stress, there was clearly no opportunity for French listeners to employ the procedure characteristic of English listening. Note, however, that no such implication holds for the use of the characteristic French procedure, syllabic segmentation, by English listeners: syllabic segmentation is a procedure potentially open to all language users, since the syllable is a unit of phonological description applicable to all languages. Nevertheless, syllable *boundaries* are not well signalled in English (or, indeed, in any stress language) , suggesting that syllabic segmentation might not prove efficient for English; indeed, explicit experimental test revealed no evidence of syllabic segmentation in English (Cutler, Mehler, Norris & Segui, 1986).

This early research thus revealed segmentation procedures which were apparently quite different being used by speakers of two languages which were historically very closely related, and indeed by the standards of world language variation must still rank as extremely close. French and English are, however, phonologically dissimilar in certain respects, and it is precisely these dissimilarities which appear to be relevant for the segmentation of spoken laguage.

The findings concerning French and English in conjunction led to the proposal for a language-universal umbrella covering the apparent language-specificity. This proposal was that speech segmentation could be based on language rhythm, and it was motivated simply by the observation that English is characterised by a stress-based rhythm and English listeners use stress-based segmentation, while French has syllabic rhythm and French listeners use syllabic segmentation. This line of argument then led in turn to the prediction that moraic structure, the basis of Japanese rhythm, would prove to be relevant for speech segmentation by Japanese listeners.

This prediction was indeed confirmed, as further research to be reported in the Forum describes (see also Otake, Hatano, Cutler & Mehler, 1993 ; Cutler & Otake, 1994). Moreover, moraic segmentation was shown to be used neither by French listeners (Otake, Hatano, Cutler & Mehler, 1993) nor by English listeners (Otake, Hatano, Cutler & Mehler, 1993 ; Cutler & Otake, 1994). Like stress-based and syllabic segmentation, moraic segmentation seemed to be a language-specific segmentation procedure. Again, phonological dissimilarities between languages were mirrored by dissimilarities in the procedures used by listeners to segment speech signals for lexical access.

These findings have very obvious implications for the learning of one language by speakers already in command of the other : application of the native segmentation procedure to the second language may prove to be inefficient. Indeed, some studies in the series produced positive

evidence that listeners *do* apply their native segmentation procedures to foreign-language input; thus French listeners were shown to apply syllabic segmentation to English (Cutler, Mehler, Norris & Segui, 1986) and to Japanese (Otake, Hatano, Cutler & Mehler, 1993), while Japanese listeners applied moraic segmentation where possible to input in English (Cutler & Otake, 1994). Research in progress (e.g. Kearns, 1994) is addressing further aspects of this question. Here, however, we are concerned less with the practical implications of the language-specifity of segmentation procedures than with defining their range and establishing the extent to which they actually function in on-line speech processing by adult listeners.

2. The Necessity of Segmentation Procedures for Adult Listening

The focus of our research is on the question of actual use by listeners of segmentation procedures of this type. Models of spoken word recognition exist which involve no explicit segmentation at all, but hold instead that word boundary information arises from the normal processes of recognising words.

There are two basic classes of such alternative models, those based on strictly sequential processing and those based on processes of competition between word candidates. The sequential segmentation models date from the 1970s (e.g. Cole & Jakimik, 1978 ; Marslen-Wilson & Welsh, 1978) and have now largely become no longer viable as a result of evidence from large vocabulary analyses made possible by current computational techniques. Their claim is that recognition of words in temporal order allows unambiguous information about each word's onset to be automatically provided by successful recognition of the preceding word. However vocabulary studies have shown that most polysyllabic words have other words embedded within them (McQueen & Cutler, 1992) and that most shorter words can be continued to form longer words (Luce, 1986); the implication is that unambiguous seg-

mentation of speech signals into words in strict temporal order is rarely possible. Cutler (1994) and McQueen, Norris and Cutler (1994) spell out this argument in greater detail.

A more serious challenge to the necessity of segmentation procedures in adult listening is provided, however, by current connectionist models of spoken word recognition such as TRACE (McClelland & Elman, 1986) and SHORTLIST (Norris, 1991, 1994). These models also avoid the need for explicit segmentation procedures, but they avoid them by postulating a process of competition between word candidates, out of which segmentation arises via the eventual success in the competition of the particular sequence of candidates which uniquely accounts for all of the input string.

Note that there is one argument in principle against any account of recognition which allows segmentation to arise from lexical processing, namely that it is dependent on the presence of a lexicon and as such fails to provide a solution to the segmentation problem which is simultaneously applicable to the case of adult and infant listeners. The infant's speech processing situation does not, in the first instance, involve recognition at all, because the infant possesses no stock of known words to recognise. Nevertheless the infant must achieve segmentation of the input in some way, in order to identify which parts of the speech signal need to be stored as units, i.e. in order to begin the process of compiling a personal lexicon. Explicit segmentation procedures such as those described in the preceding section offer a solution to the segmentation problem which is in principle accessible to the prelinguistic infant, and they are therefore attractive for the very reason that they would provide a unified solution to the segmentation problem as experienced by the infant and the adult (see Cutler, 1994: Christophe, 1993 for further consideration of this point).

Nevertheless, it is obviously a logical possibility that no such unified situation exists. The infant situation is a special case and it may well

be that once the initial segmentation problem has been solved and a sufficient lexical stock has been assembled, processes of explicit segmentation are rendered unnecessary because the normal processes of recognition achieve segmentation without them. Thus explicit segmentation would be used by the infant, but adult recognition would involve only lexically driven processes ; the research described in other papers in the Forum and outlined briefly in section 1 might then be held to show either that adult listeners *can* invoke explicit segmentation procedures if they have to, or it might be, in the worst case, that these results could also be accounted for by lexical processes.

Indeed, such an alternative interpretation might in principle be offered for one of the basic results for English, namely that of Cutler and Norris (1988). Their finding was that CVCC words such as *mint* are harder to recognise if they are embedded in nonsense bisyllables such as *mintayf* than if they are embedded in bisyllables like *mintef.* The crucial difference between these two stimuli is that in the first the vowel following the embedded word is strong ([e]), whereas in the second it is weak (schwa). Cutler and Norris argued that stress-based segmentation processes would take any syllable containing a strong vowel as the likely beginning of a new word, so that *mintayf* would be segmented *min-tayf,* rendering detection of *mint* more difficult because the phonetic string corresponding to the word was interrupted by a segmentation point. In *mintef* no such segmentation, and hence no inhibition of detection of *mint,* would arise.

However, the reason that stress-based segmentation - a procedure of assuming that strong syllables are likely to be word-initial-works effectively for English is that most strong syllables *are* indeed word-initial, as Cutler and Carter (1987) showed. The English vocabulary contains many more words beginning with strong than with weak syllables. One could argue, therefore, that the difficulty of detecting *mint* in *mintayf* in comparison with *mintef* might arise from the presence of a larger

number of alternative candidate words competing for the second sylla-
ble in the first compared with the second case. Thus competition-based
models might be held to account for findings apparently supporting
explicit segmentation procedures. In the next section we describe a
series of experiments which we undertook in English testing the predic-
tions both of competition models and stress-based segmentation models.

3. Competition and Segmentation in Adult Listening

Our experiments all used the word-spotting task developed by Cutler
and Norris (1988). In this task listeners hear nonsense utterances
(which were all isolated bisyllables in these studies as in the earlier
one), and are required to press a button whenever they detect a real
word in the input. They do not know what words might occur; their
task is simply to respond to any real word they hear. When they do
detect a real word, they then have to repeat it (and of course if a
listener says the wrong word on any trial, the corresponding response
is not included in the reaction time analysis). Both response time and
miss rate can be used as measures of word recognition difficulty. In the
Cutler and Norris (1988) study, as described above, word recognition
difficulty was found to vary as a function of the phonological structure
of the utterances: CVCC words in CVCCVC utterances were more
difficult to detect if the second vowel was strong than if it was weak.
This result was taken as evidence of an explicit stress-based segmenta-
tion procedure.

All experiments of the current study investigated effects of inter-
word competition and effects of explicit segmentation procedures in
parallel. In Experiments 1-3 (which are reported in full detail in
McQueen, Norris & Cutler, 1994), we tested competition effects by
manipulating whether or not the matrix utterance (i.e. the nonsense
utterance in which the target word occurred) could itself be continued
to form another real word. For instance, [dames] is a nonsense bisylla-

ble in itself, but it has the real word *mess* contained within it, and it is also the first two syllables of another real word, *domestic.* Likewise, [saekraf] is a nonsense bisyllable which contains the real word *sack* and could be continued to form another real word, *sacrifice.*

If competition between word candidates is occurring during word recognition, then the utterance [demes] should cause competition between *mess* and *domestic,* and [saekraf] should cause competition between *sack* and *sacrifice.* The monosyllabic target words should be harder to detect when such competition is occurring than when there is no competition. Therefore we compared detection of words like *mess* and *sack* in the competition contexts [dames] and [saekref] and in non-competition contexts such as [names] and [saekrek]. The latter are non-competition contexts because neither of them can be continued to form a real word.

At the same time we conducted a further test of stress-based segmentation procedures. Cutler and Norris (1988) proposed that listeners segment speech at the onset of every strong syllable. This means that a real word consisting of a single strong syllable (such as *mess* or *sack)* should be segmented from any syllable that precedes it, but should not be segmented from any weak syllable that follows it. Thus such words should be easier to detect in weak-strong (WS) utterances than in strong-weak (SW); in the above examples, it should be easier to detect *mess* in its two WS contexts than *sack* in its two SW contexts. Of course, that particular comparison could be confounded by effects of competition and by simple differences in detectability between words like *mess* and words like *sack.* So we instituted a more rigorous test of stress-based segmentation by comparing recognition of each word in its non-competition context with recognition of the same word in another non-competition context, with the opposite stress pattern. Thus detection of *mess* in WS [names] was compared with detection of *mess* in SW [mestem], and detection of *sack* in SW [saekrek] was compared

with detection of *sack* in WS [klasaek].

Experiment 1 produced strong effects of both competition and stress-based segmentation. The competition effect showed itself in the comparison between contexts: *mess* was detected both more rapidly and more accurately in the non-competition context [names] than in the competition context [dames], and *sack* was detected more accurately (though not more rapidly) in the non-competition context [saekrek] than in the competition context [saekraf]. The effect of stress-based segmentation showed itself in the comparison between the two stress patterns in the non-competition contexts: *mess* and *sack* were detected more rapidly and more accurately in the WS patterns [names] and [klasaek] than in the SW patterns [mestem] and [saekrek].

Thus in Experiment 1 stress-based segmentation and inter-word competition appeared to be operating simultaneously. In Experiment 2 we made the listeners' word-spotting task a little easier by constraining where in the utterance the words were to be spotted; half of the subjects listened for words occurring at the beginning of the nonsense bisyllables, half of the subjects listened for words cccurring at the bisyllables' ends. Exactly the same effect of stress-based segmentation appeared: the words were detected faster and more accurately in WS than in SW contexts (this comparison was in this case between separate groups of subjects, of course). There was also a strong effect of competition for the WS utterances: it was still harder to detect *mess* in [dames] than in [names]. Subjects who were listening for words at the ends of the bisyllables of course had to listen to the whole bisyllable, so *domestic* had the same opportunity to compete with *mess* as in Experiment 1. However there was no effect of competition for the SW words in Experiment 2 : *sack* was as easy to detect in [saekraf] as in [saekrek], This also makes good sense, because in this case the subjects detecting the words in SW bisyllables were instructed to attend only to the beginnings of the utterances. Once they had spotted *sack* at the begin-

ning of either of the items in which it occurred, there was no reason to attend to the second syllable, and it was only in the second syllable that *sacrifice* stopped competing in [saekrek] but continued competing in [saekref].

In Experiment 3 the same bisyllables were presented to a new group of listeners, but they were digitally expanded or compressed so that the embedded monosyllabic words were roughly equal in duration in SW and WS bisyllables. This is because the strong syllables were longer in the original WS utterances than in the SW, and the added duration could have been the source of the apparent stress-based segmentation effect, i. e. the advantage for words in WS over SW patterns. But this was not the case: the same effects of stress-based segmentation as seen in Experiments 1 and 2 reappeared in Experiment 3, in both response time and miss rates. Also there were again strong effects of competition (for the words in WS bisyllables).

In Experiment 4 (reported in full in Norris, McQueen & Cutler, in press) we adopted a different approach to testing effects of competition. In this study we manipulated the actual number of competing words. To do this we used a large computer-readable dictionary (the *Longmans Dictionary of Contemporary English)* to establish CV sequences which constituted the onset of many versus few words in the English language. Then these sequences were built into CVCCVC bisyllables, in the second CV position. The target word was a CVCC monosyllable. The final phoneme of the embedded word and its following vowel therefore constituted the onset of many versus few words of English.

For instance, there are many English words beginning with [kʌ] *(cup, couple, custard, cudgel, culinary, cover, cuddle, company,* and so on and on), just as there are many words beginning with *[ke] (canoe, collide, connect, cavort, commercial, corrosion, cathedral, cadet,* and so on). Thus *mask* in both [maskʌk] and [maskek] will have many potential words beginning from its final phoneme. On the other hand,

there are few words beginning [tau] (four, in fact: *town, towel, tout* and *tousle,* plus their morphological relatives). Likewise, there are few words beginning *[te] (terrain, toboggan, telephony* etc.). Thus *mint* in [mintaup] and [mintep] will have few potential words beginning from its final phoneme.

Each CVCC target word had a matched CVC target word, in the initial position in a CVCCVC context, as a control. Thus *mask* was compared with *pass* (in [paskAk] and [paskek]), while *mint* was compared with *thin* (in [0intaup] and [0intep]).

In this fourth word-spotting experiment, then, a competition effect would show itself in a difference between items with many versus few potential words beginning with the second CV in the CVCCVC string. The effect should be expected to be different for CVCC and CVC target words. This is because the crucial second CV overlaps with the final phoneme of *CVCC* words, so the availability of many competitors might hinder recognition of the target word. However the second CV has no overlap with the CVC target words, so it is possible that the availability of many competitors might actually facilitate recognition of the target word, by emphasising the segmentation point.

Stress-based segmentation should show itself in the same way as in the Cutler and Norris (1988) study: CVCC words should be harder to detect when the following vowel is strong than when it is weak. No effect of the following vowel should be observed with CVC target words (again, this replicates a finding of Cutler and Norris).

The results of this study again showed effects both of stress-based segmentation and of inter-word competition. The segmentation effect showed itself in a replication of Cutler and Norris' findings: CVCC words were detected both more rapidly and more accurately in SW than in SS bisyllables (i.e. *mask* was easier to detect in [maskek] than in [maskAk], and *mint* was easier to detect in [mintep] than in [mintaup]). No such effect of the second vowel was observed with CVC

target words.

The competition effect showed itself mainly with the CVC words, which were much easier to detect when there were many words beginning with the CV sequence following the target *(pass* in [paskAk] and [paskek]) than when there were few *(thin* in [0intaup] and [0intep]). This implies that, as expected, the availability of competitors for the second syllable makes the first syllable easier to recognise on its own. In the CVCC words, the competition effect showed itself indirectly: the segmentation effect (advantage of words in SW versus SS bisyllables) was larger in response time when there were many competitors *(mask)* than when there were few *(mint),* and moreover there was a significant segmentation effect in miss rates (fewer targets missed in SW than in SS bisyllables) only when there were many competitors, not when there were few. This implies that when stress-based segmentation has occurred (in SS bisyllables), more competitors exert a stronger pull on the final consonant of the embedded word than few competitors do.

Independently of our studies, Vroomen & de Gelder (1995) also demonstrated an effect of number of competitors in a priming/segmentation task. Dutch CVCC words embedded in the initial portion of CVCCVC contexts produced less priming (i. e. were less highly activated) when there were many potential words beginning from, and hence competing for, the second medial consonant (analogous to our [maskAk] case) than when there were few (analogous to our [mintaup] case).

The results of our Experiment 4, however, go beyond demonstrating that number of competitors exerts an effect in word recognition. Firstly, they show that the stress-based segmentation effect is present irrespective of the number of competitors available - in other words, the results of Cutler and Norris (1988) cannot be ascribed to effects of competition. Secondly, these results show that lexical competition and segmentation effects interact in that increased competitor availability

can make the segmentation effect larger.

4. Modelling Segmentation with Competition

The results of our experiments have shown that both inter-word competition and explicit segmentation play a role in the recognition of spoken words. Note that this is the first direct and unequivocal experimental evidence for competition effects. Many previous studies had shown that words may be concurrently activated when they overlap in spoken input (e.g. Taft, 1986; Goldinger, Luce & Pisoni, 1989; Zwitserlood, 1989; Cluff & Luce, 1990; Shillcock, 1990; Goldinger, Luce, Pisoni & Marcario, 1992). However, concurrent activation does not necessarily imply active competition. To demonstrate competition, it is necessary to show that other competing words may inhibit recognition of a given word. Exactly such evidence of inhibition has been provided in the present studies.

Models of spoken word recognition involving competition have been implemented as connectionist programs. TRACE (McClelland & Elman, 1986) and SHORTLIST (Norris, 1991, 1994) are two such fully worked out models. In their initial formulations, neither model involves explicit segmentation. Both would allow segmentation simply to fall out of the process of inter-word competition. However the present results show that in this respect pure competition models are inadequate as a representation of human word recognition processes. Human listeners do use competition, but they also use explicit segmentation. Can explicit segmentation be incorporated into a competition-based model of word recognition ?

We have successfully modified SHORTLIST to incorporate a process of explicit segmentation. Note that it would be rather difficult to simulate our present results with TRACE, for practical reasons; TRACE'S speed of operation is highly dependent on vocabulary size, and hence TRACE simulations are usually conducted with only tiny

vocabularies of a few hundred words. (This is an artefact of the solution which McClelland and Elman adopted to simulate the temporal nature of speech — TRACE'S entire vocabulary is replicated at each time slice of the input. This implausible architecture does work to simulate competition, but it obviously makes the use of large vocabularies computationally very expensive.) Since our calculations of competitor effects were based on a real dictionary, to simulate our results in TRACE it would be necessary to tailor TRACE'S vocabulary to mirror exactly the proportional distributions in the vocabulary as a whole.

SHORTLIST is a hybrid model which runs effectively on a realistically sized vocabulary — indeed, it runs on the *Longman Dictionary of Contemporary English* (26000 words in the version used in our studies). The initial stage of the model is entirely driven by the acoustic/phonetic properties of the input. This initial stage creates a short set of candidate words consistent in whole or in part with the input (the *shortlist)* ; these words are then wired into a network and allowed to compete among themselves until a winner emerges.

Table I.

Position :	-1	C	$+1$	$+2$	$+3$	$+4$
Weak-Strong						
mess in [dəmɛs]	0.003	0.148	0.293	0.306	0.306	0.306
mess in [nəmɛs]	0.011	0.202	0.311	0.311	0.311	0.311
sack in [kłəsæk]	0.016	0.207	0.307	0.307	0.307	0.307
domestic in [dəmɛs]	0.137	0.263	0.021	-0.103	-0.144	-0.157
Strong-Weak						
sack in [sækrəf]	0.088	0.210	0.184	0.168	0.222	0.290

sack in [sækrək]	0.088	0.210	0.191	0.244	0.308	0.308
mess in [mɛstəm]	0.080	0.249	0.269	0.281	0.309	0.310
sacrifice in [sækrəf]		−0.009	0.072	0.212	0.116	−0.057
sacrifice in [sækrək]		−0.009	0.041	−0.027	−0.109	−0.151

Table 1. Mean activation values of the target words and the embedding words from Experiments 1-3, over time, in SHORTLIST. Values are given for both weak-strong and strong-weak strings. In the weak-strong utterances, mean activations are given for targets embedded in competition (e.g. *mess* in [dames], the onset of *domestic),* and in non-competition contexts (e.g. *mess* in [names], *sack* in [klesaek]). Also given are the activation values of the embedding words in the competition contexts (e.g. *domestic* in [demes]). In the strong-weak strings, values are again given for targets embedded in competition (e.g. *sack* in [saekref], the onset of *sacrifice)* and non-competition contexts (e.g. *sack* in [saekrak], *mess* in [mestem]). Also shown are the values for embedding words both in competition (e.g. *sacrifice* in [saekraf]) and non-competition contexts (e.g. *sacrifice* in [saekrek]). The values are aligned with the last consonant of the target word ("C"). Positions before C are for each phoneme working back through each item; positions after C are for following segments or silence markers.

Table 1 shows how SHORTLIST models the recognition of the target words used in Experiments 1 to 3. Activation values are shown for successive points representing time (in phoneme-sized slices), where the point labelled C represents the final consonant of the target word. In the WS utterances the points following C represent silence. It can be seen that the activation of *mess* is lower in [dem3s] than in [nam3s] at the final consonant (because of competing activation from *domestic)* ; *mess* in [dem3s] does not recover until silent slices of time

have occurred and effectively removed *domestic* from competition. In the SW utterances the activation of the target word (e.g. *sack)* at C is equivalent in competition ([saekref]) and non-competition contexts ([saekrek]), but it rises when subsequent phonetic information removes the competitor in the non-competition context. In the competition context activation of the target word only rises later, when silence has removed the competitor from that context as well (note that these values are averaged over many items ; in most cases the disambiguating phonetic information arrived at the second phoneme after the target word although in a few cases, such as [saekref/saekrek], it arrived at the third).

   Thus SHORTLIST very accurately models the competition effects which we observed in Experiments 1 to 3. But the segmentation effects that we observed there are not directly captured in this simulation. SHORTLIST in this unaugmented version also does not capture the segmentation effects we observed in Experiment 4, which amounted to a more rigorous test of segmentation, with competition controlled. Table 2 shows the relevant activation values for the CVCC target words; it can be seen that although the targets have lower activation values at and beyond C when there are many competitors than when there are few, activations do not differ as a function of whether the following vowel is strong or weak.

Table 2.

Position :	-2	-1	C	$+1$	$+2$	$+3$
Many competitors						
mask in [maskʌk]	- 0.053	0.196	0.262	0.288	0.556	0.615
mask in [maskək]	- 0.053	0.196	0.262	0.275	0.507	0.509
Few competitors						
mint in [mɪntaup]	- 0.048	0.152	0.389	0.600	0.620	0.620

| *mint* in [mɪntəp] | − 0.048 | 0.152 | 0.389 | 0.492 | 0.517 | 0.543 |

Table 2. Mean activation values of the CVCC target words from Experiment 4, over time, in SHORTLIST, with no segmentation procedure implemented. Values are shown for targets in strong-strong strings and strong-weak strings, in which there were either many competitor words beginning from the last consonant of the target (e.g. *mask* in [maskAk] and [maskek], with many words beginning from the medial [k]) or few competitor words (e.g. *mint* in [mintaup] and [mintep], with few words beginning from the [t]). The values are aligned with the last consonant of the target word ("C" ; e.g. the [t] in *mint).* Positions before C are for each phoneme working back through each item; positions after C are for following segments or silence markers.

To incorporate the effects of stress-based segmentation in SHORT-LIST, we added two features to the model. Firstly, we instituted a penalty on lexical candidates containing no strong syllable onset where there is one in the input. This mechanism was intended to simulate Cutler and Norris' stress-based segmentation proposal in that it stipulates that strong syllable onsets are to be viewed as segmentation points. Secondly, we provided an activation boost to all word candidates beginning with a strong syllable. This simulates the aspect of Cutler and Norris' stress-based segmentation proposal which proposes that the purpose of segmentation is to initiate lexical access attempts from strong syllables.

Table 3.

Position :	−2	−1	C	+1	+2	+3
Many competitors						
mask in [maskʌk]	− 0.053	0.196	0.213	0.001	0.337	0.499
mask in [maskək]	− 0.053	0.196	0.262	0.275	0.507	0.509
Few competitors						
mint in [mɪntaup]	− 0.048	0.152	0.324	0.350	0.505	0.505
mint in [mɪntəp]	− 0.048	0.152	0.389	0.492	0.517	0.543

Table 3. Mean activation values of the target words from Experiment 4, over time, in SHORTLIST, with the stress-based segmentation procedure implemented with a combined penalty and boost (see text for details) .Values are shown for targets in strong-strong strings and strong-weak strings, in which there were either many competitor words beginning from the last consonant of the target (e.g. *mask* in [maskAk] and [maskek], with many words beginning from the medial [k]) or few competitor words (e.g. *mint* in [mintaup] and [mintep], with few words beginning from the [t]). The values are aligned with the last consonant of the target word ("C"; e.g. the [t] in *mint).* Positions before C are for each phoneme working back through each item; positions after C are for following segments or silence markers.

Table 3 shows the same items as in Table 2, as modelled by the augmented version of SHORTLIST. Now it can be seen that activation of the target words is consistently higher, at and beyond C, when the following vowel is weak than when it is strong. Moreover, the difference is much greater when there are many competitors than where there are few. This is exactly the pattern of results observed in Experiment 4. Thus the augmented version of SHORTLIST correctly captures both the presence of intei-word competition and the use of

explicit stress-based segmentation.

5. Conclusion

The viability of explicit segmentation procedures in human speech recognition is not compromised by the power of inter-word competition as a principle of word recognition. Competition could in principle make explicit segmentation unnecessary. And competition does indeed play a role in word recognition by human listeners, as our experiments, and those of others, have shown. But listeners do not rely on competition alone: they also make use of explicit segmentation procedures - specifically, the English listeners in our studies make use of stress-based segmentation.

We have shown that explicit segmentation procedures can be incorporated in a competition-based model of word recognition, and that the model then simulates the experimental findings very accurately. Note that there is no reason to believe that this solution would not also work in principle for other languages. We happen to have implemented SHORTLIST with an English dictionary, and we have augmented it with the stress-based segmentation procedure characteristic of English listeners. But SHORTLIST is not itself a model of English word recognition; it is a universal model of human word recognition. Thus extension of the present technique to the modelling of word recognition in other languages is quite straightforward. The model would work in just the same manner if given a dictionary of French and input in French, or a dictionary of Japanese and input in Japanese, and so on: the input would serve to activate a shortlist of potential word candidates, which would then compete among themselves for recognition. Likewise, the segmentation procedures used by listeners in other languages lend themselves as well to incorporation in SHORTLIST as stress-based segmentation does. Thus the syllable-based segmentation procedure characteristic of French listeners, and the mora-based seg-

mentation procedure characteristic of Japanese listeners, could each be implemented in an augmented version of SHORTLIST, with an appropriate dictionary, and applied to speech input in the relevant language.

The modelling of lexical access from continuous speech input has made great progress in recent years. Competition techniques are a very powerful addition to the modelling repertoire. Moreover, empirical findings suggest that they give a true picture of human word recognition. They exist in conjunction, however, with techniques of explicit speech segmentation, for which there is also abundant empirical evidence. Such segmentation procedures are language-specific ; this implies that correct universal modelling of lexical access will require language-specific adjustments to word recognition models.

Acknowledgements

References

Christophe, A. (1993). *Role de la prosodie dans la segmentation en mots.* (PhD Dissertation, Ecole des Hautes Etudes en Sciences Sociales, Paris)

Guff, M. S. & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology : Human Perception & Performance,* 16, 551-563.

Cole, R. A. & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of Information Processing* (pp. 67-116). London: Academic Press.

Cutler, A. (1994). Segmentation problems, rhythmic solutions. *Lingua,* 92, (in press).

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation : Evidence from juncture misperception. *Journal of Memory & Language,* 31, 218-236.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language,* 2, 133-142.

Cutler, A., Mehler, J., Norris, D. & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology,* 24, 381-410.

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

Cutler, A. & Otake, T. (in press). Mora or phoneme ? Further evidence for language-specific listening. *Journal of Memory & Language,* 33.

Goldinger, S. D., Luce, P. A. & Pispni, D. B. (1989). Priming lexical neighbours of spoken words: Effects of competition and inhibition. *Journal of Memory & Language,* 28, 501-518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B. & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology : Learning, Memory & Cognition,* 18, 1211-1238.

Kearns, R. K. (1994). *Prelexical speech processing in mono- & bilinguals.* PhD thesis, University of Cambridge.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics,* 39, 155-158.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech.

Cognitive Psychology, 10, 29-63.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology,* 18, 1-86.

McQueen, J. M. & Cutler, A. (1992). Words within words: Lexical statistics and lexical access. *Proceedings of the Second International Conference on Spoken Language Processing,* Banff, Canada; Vol. 1, 221-224.

McQueen, J. M., Norris, D. G. & Cutler, A. (1994). Competition in spoken word recognition : Spotting words in other words. *Journal of Experimental Psychology : Learning, Memory & Cognition,* 20, 621-638.

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

Norris, D. G (1991). Rewiring lexical networks on the fly. *Proceedings of EUROSPEECH* 91, Genoa; Vol. 1, 117-120.

Norris, D. G (1994). SHORTLIST : A connectionist model of continuous speech recognition. *Cognition,* 52, 189-234.

Norris, D. G., McQueen, J. M. & Cutler, A. (in press). Competition and segmentation in spoken word recognition, *Journal of Experimental Psychology : Learning, Memory & Cognition,*

Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory & Language,* 32, 358-378.

Segui, J., Frauenfelder, U. H. & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology,* 72, 471-477.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 24-49). Cambridge MA: MIT Press.

Taft, M. (1986). Lexical access codes in visual and auditory word

recognition. *Language & Cognitive Processes,* 1, 297-308.

Vroomen, J. & de Gelder, B. (1995) Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology : Human Perception & Performance,* 21.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition,* 32, 25-64.