

# The Possible-Word Constraint in the Segmentation of Continuous Speech

Dennis Norris

*MRC Applied Psychology Unit, Cambridge, United Kingdom*

James M. McQueen and Anne Cutler

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

and

Sally Butterfield

*MRC Applied Psychology Unit, Cambridge, United Kingdom*

We propose that word recognition in continuous speech is subject to constraints on what may constitute a viable word of the language. This Possible-Word Constraint (PWC) reduces activation of candidate words if their recognition would imply word status for adjacent input which could not be a word - for instance, a single consonant. In two word-spotting experiments, listeners found it much harder to detect *apple*, for example, in *fapple* (where [f] alone would be an impossible word), than in *vuffapple* (where *vuff* could be a word of English). We demonstrate that the PWC can readily be implemented in a competition-based model of continuous speech recognition, as a constraint on the process of competition between candidate words; where a stretch of speech between a candidate word and a (known or likely) word boundary is not a possible word, activation of the candidate word is reduced. This implementation accurately simulates both the present results and data from a range of earlier studies of speech segmentation. © 1997 Academic Press

## INTRODUCTION

When we listen to continuous speech, we apprehend a sequence of discrete words. The effortlessness with which we understand speech might suggest that recognition of spoken words is a rather simple task; but in fact it is a complex process, in which a very variable and largely continuous speech signal has to be mapped onto a lexicon of tens of thousands of words. The

We thank the Longman Group, U.K. Inc. for allowing us to use a machine-readable version of the *Longman Dictionary of Contemporary English* (Procter, 1975).

Address correspondence and reprint requests to Dennis Norris, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK. Fax: (44) 1223 359062. E-mail: dennis.norris@mrc-apu.cam.ac.uk.

problem of spoken word recognition is already severe when one considers isolated words. The listener must be able to select, in spite of variability in the speaking rate, dialect, and other voice characteristics of a speaker, and in spite of any background noise, the word the speaker intended. Further, this word must be selected from a lexicon in which many words differ from one another very little, often by no more than one distinctive feature.

The problem becomes even more severe, however, in normal spoken language—continuous speech. The difficulties of selecting a particular word for a given stretch of continuous input may appear to be analogous to those for isolated word recognition, but there is a catch: the listener has no way of knowing, a priori, where in the signal a new word has begun, and thus cannot assume that a particular portion of the input corresponds to an individual word. That stretch of speech may just as well map onto two words. For example, the string [lɛtəs] embedded in continuous speech may be the word *lettuce*, but it could just as easily be the two words *let* and *us*, or it might form part of some longer sequence of words, such as *his roulette astounds me*. The central problem for continuous speech recognition lies in the fact that there are no fully reliable cues to word boundaries (Lehiste, 1972; Nakatani & Dukes, 1977), and thus recognition also necessitates a process whereby the input is segmented into individual words.

How is this segmentation problem solved by human listeners? A great deal of research has been devoted to this topic in recent years, and from this research two general conclusions emerge: first, listeners are capable of exploiting a range of cues in the signal, some direct and some indirect, even where the information thus obtained is probabilistic rather than deterministic, and second, listeners entertain multiple hypotheses about the input which are simultaneously active and effectively in competition with one another. The evidence on these issues is summarized in the following sections.

### SEGMENTATION CUES IN THE INPUT

Physical cues to word boundaries in continuous speech do exist; they include lengthening of onset syllables and segments (Gow and Gordon, 1995; Lehiste, 1972), lengthening of final syllables (Beckman & Edwards, 1990), and aspiration of word-initial stops (in English, Lehiste, 1960; Trager & Bloch, 1941). Phonotactic cues also exist; some sequences of segments (such as [mr] in many languages) cannot co-occur within a syllable, and therefore signal likely word-boundary locations. These cues are, as already noted, not consistent; many word boundaries are not marked in this way. Nevertheless, as argued by Church (1987), an efficient system should be able to make use of such information when it is available. In off-line tasks, such as choosing between alternative interpretations of a segmentally ambiguous string like *nitrate/night rate*, listeners can exploit the physical cues which disambiguate such strings (Christie, 1974; Quené, 1989, for Dutch). Recent evidence has shown that phonotactic cues can be exploited in on-line segmentation. McQueen (submitted), using the word-spotting task, in which listeners are

required to detect any real word embedded in a nonsense input, showed that words are very difficult to spot when they are misaligned with a phonotactically determined syllable boundary. Dutch listeners were presented with bisyllabic nonsense strings such as [fi.drɔk] and [fim.rɔk] (the period in the phonetic transcription indicates the location of a mandatory syllable boundary). Both strings contain the Dutch word *rok* (skirt), but in the former this embedded word is misaligned with a syllable boundary which is mandatory on phonotactic grounds ([d] cannot be syllable-final), while in the latter the word is exactly aligned with a mandatory boundary (a [mr] cluster is illegal). Detection of the word was significantly slower and less accurate in [fi.drɔk] than in [fim.rɔk].

Similarly, in some languages phonological restrictions on the segmental content of words can be exploited in segmentation. Finnish, for instance, manifests a vowel harmony constraint, according to which any of the three back vowels /a/, /o/, /u/ may not co-occur within a word with any of the three front vowels /æ/, /ø/, /y/. Again, this cue cannot determine all word boundaries; there are two additional "neutral" vowels which may co-occur with either class, and of course two successive words in speech may contain vowels of the same class. Nevertheless, in the case that two successive syllables in a spoken utterance contain vowels from the two mutually exclusive classes, there must be a word boundary between the syllables; as pointed out by Trubetzkoy (1939), this cue could be valuable to listeners in solving the segmentation problem. Suomi, McQueen, and Cutler (1997) showed that Finnish listeners do exploit such information on-line; words such as *käry* (odor) were significantly easier to detect in a nonsense string such as *pokäry* (in which the preceding context syllable contains a vowel from the back harmony class and thus mismatches with the front vowels in *käry*) than in *pökäry*, with a context syllable containing a vowel from the same class as the vowels of the embedded word.

The metrical structure of speech provides listeners with a further source of information for segmentation. Because languages differ rhythmically, the same kind of metrical information cannot be used across all languages; nevertheless, evidence from many languages shows that metrical structure is used for segmentation. In French, Catalan, and Spanish, for example, the syllable is the basic metrical unit, and native speakers of these languages use syllabic information in segmentation (Cutler, Mehler, Norris, & Seguí, 1986, 1992; Mehler, Dommergues, Frauenfelder, & Seguí, 1981; Pallier, Sebastián-Gallés, Felguera, Christophe, & Mehler, 1993; Sebastián-Gallés, Dupoux, Seguí, & Mehler, 1992). In Japanese, the basic metrical unit is the mora, and accordingly it appears that Japanese listeners use moraic information in segmentation (Cutler & Otake, 1994; Otake, Hatano, Cutler & Mehler, 1993). Speakers of stress-timed languages like English and Dutch use the rhythmic distinction between strong and weak syllables for segmentation (Cutler & Butterfield, 1992; Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995; Vroomen & de Gelder, 1995; Vroomen, van

Zon, & de Gelder, 1996). Strong syllables contain full vowels, weak syllables reduced vowels; in English, this distinction of vowel quality outweighs a contrast between stressed versus unstressed syllables (Fear, Cutler, & Butterfield, 1995).

The metrical segmentation procedure in English was originally formulated as a strategy: the speech signal should be segmented at strong syllables, and lexical access attempts should be initiated at these segmentation points (Cutler & Carter, 1987; Cutler & Norris, 1988). This Metrical Segmentation Strategy (MSS) would work well for English, since over 90% of content words begin with strong syllables, and about 75% of all strong syllables in a corpus of English speech were at word onsets (Cutler & Carter, 1987). The MSS has also received considerable experimental support. Cutler and Butterfield (1992) reported that in both spontaneous and laboratory-induced slips of the ear, listeners tend to insert word boundaries before strong syllables and to delete them before weak syllables. Cutler and Norris (1988) found that words were more difficult to spot at the beginning of bisyllabic nonsense strings when the second syllable was strong (such as *mint* in [mɪntɛɪf]) than when the second syllable was weak (*mint* in [mɪntəf]). According to the MSS, segmentation occurs at the second strong syllable in [mɪntɛɪf], so *mint* must in effect be reconstituted over this segmentation point, making detection more difficult. Similar results, in both boundary misperceptions and in word spotting, have been obtained in Dutch (Vroomen et al., 1996). The lexical statistics of Dutch are also favorable for the MSS—about 85% of Dutch words begin with strong syllables (Schreuder & Baayen, 1994; Vroomen & de Gelder, 1995).

Listeners therefore have a number of means at their disposal to exploit the phonological characteristics of the input to select the most likely locations for word boundaries. There are, however, quite independent characteristics of the word recognition process itself, as the next section outlines, which act to facilitate the operation of segmentation.

### SEGMENTATION VIA COMPETITION

A spoken input activates multiple candidates with which it is fully or partially compatible, and these candidates for recognition compete among one another. A great deal of experimental evidence now underpins this conclusion. Studies of the recognition of isolated monosyllabic words, for example (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Luce, Pisoni, & Goldinger, 1990; Slowiaczek & Hamburger, 1992), have shown that recognition can be delayed when a phonetically similar prime word precedes the target. These results have been interpreted as evidence that both prime and target are activated when the prime word is heard; competition between these words causes the target word to be inhibited, and this inhibition persists to make recognition of the target, when it is subsequently presented, more difficult. Cluff and Luce (1990) also found effects of phonetic similarity suggesting competition between multiple candidate words. Bisyllabic words

like *madcap* were presented for identification in white noise. The number and frequency of occurrence of words similar to the first syllable (e.g., *mad*) influenced recognition performance, as did the number and frequency of words similar to the second syllable (e.g., *cap*).

Research using the cross-modal priming technique has indicated that multiple candidate words are activated during recognition (Marslen-Wilson, 1987, 1990; Shillcock, 1990; Swinney, 1981; Tabossi, Burani & Scott, 1995; Zwitserlood, 1989). Zwitserlood (1989), for example, found that when Dutch listeners hear the string [kapit], which is consistent with both *kapitein* (captain) and *kapitaal* (capital), lexical decision on a visual probe is faster for probes which are semantic relatives of either word (e.g., *schip*, ship, and *geld*, money), suggesting that both *kapitein* and *kapitaal* were activated. Similarly, but for a word embedded at the end of a longer word, Shillcock (1990) reported that decisions to *rib*, for example, are speeded just after listeners have heard *trombone*, presumably due to activation of *bone*. Likewise, Gow and Gordon (1995) found that listeners who heard sentences containing word sequences such as *two lips* which could also be a single word (*tulips*) responded faster both to an associate of the second word (*kiss*) and to an associate of the longer word (*flower*).

Note that Gow and Gordon (1995) did not obtain priming on *kiss* when listeners heard a sentence containing *tulips*. This last result conflicts with the result reported by Shillcock (1990). Gow and Gordon (1995) argued that acoustic cues to word boundaries, such as lengthening of onset segments, may be absent at the beginning of the second syllable of *tulips*, discouraging the activation of the embedded word. With the cross-modal priming paradigm, however, conclusions about parallel activation of lexical hypotheses can only be drawn when priming has been observed. When no priming is observed, as when *kiss* is not primed by *tulips*, nothing can be said about the degree of activation of the embedded word *lips* other than that it is not sufficient to produce a priming effect. Perhaps *lips* is simply not activated in this situation; but it is also possible that *lips* is weakly activated, but that due to the influence of number of competitors, word frequency effects, and so on, it has not reached a high enough level of activation to facilitate recognition of *kiss*. And finally, even when cross-modal priming is observed, it confirms that multiple activation has occurred but not that competition takes place between the activated hypotheses.

Cross-modal priming has also shown that word hypotheses are considered even when they span a word boundary in the input. Tabossi et al. (1995) presented Italian listeners with sentences including sequences of words such as *visi tediati* ("faces bored"). Responses to visual probes (here, *parenti* "relatives," aligned with the [d] of *tediati*) were faster in this condition than in a control condition, suggesting that the hypothesis *visite* ("visits," an associate of *parenti*) was being considered even well after the boundary between *visi* and *tediate*. This result further supports the claim that multiple lexical hypotheses are activated during recognition, and suggests that consid-

eration of these hypotheses can be long-lived, although it provides no direct evidence of competition among hypotheses.

Direct evidence for inter-word competition was provided by McQueen et al. (1994) using the word-spotting task. Listeners were presented with bisyllabic nonsense strings, some of which contained embedded words. Their task was to attempt to spot these embedded words. Detection was slower and less accurate when the nonsense string was the beginning of a longer real word than when the string could not be continued to form a real word. For example, *mess*, embedded in the second syllables of the strings [dəmɛs] and [nəmɛs], was spotted more slowly and less frequently in the first string (which can be continued to form the word *domestic*). Similarly, *sack* was spotted less accurately in the string [sækrəf] (which can be continued to form the word *sacrifice*) than in the string [sækrək]. The longer words (*domestic* and *sacrifice*) compete with the embedded words, making their detection more difficult.

This empirically demonstrated competition between simultaneously activated word candidates can in fact effectively produce segmentation of the input. For instance, consider the listener's task when presented with the spoken string *ship inquiry*, spoken by a British English speaker. Homorganic assimilation of the nasal consonant at the end of the first syllable of *inquiry* will produce a velar nasal, so that the first two syllables of the input will thus be perfectly consistent with the word *shipping*; *shipping* will accordingly at that point be the most highly activated candidate word, as it has greater bottom-up support (more phonemes) than *ship*, which is of course also activated. Subsequent incoming input will activate, for instance, *choir*. Note that no word boundaries have been signaled on the basis of explicit cues in the input, but boundaries are nevertheless being effectively postulated by the activation process, simply depending on where in the signal different candidate words would begin and end. Words straddling any possible boundary may compete with words which respect the boundary. Thus *shipping* and *choir* will compete with *inquire* and *inquiry*. Once the final segment of the string arrives, however, the candidate *inquiry* would have added support and would gain enough activation to dominate *choir*; it can then join forces with *ship* to compete strongly with *shipping*. Competition would thus finally result in a new best (and now correct) interpretation, now with *ship* and *inquiry* dominating the activation pattern, with a clear word boundary between them.

Note that one might propose that the metrical effects described in the preceding section simply arise as a consequence of lexical competition: since more English words begin with strong syllables than with weak syllables, the strong-weak differences observed could be due to asymmetric competition effects. However, McQueen et al. (1994) and Norris et al. (1995) found that metrical effects do not reduce to competition effects. Separately from the competition effects described above, McQueen et al. (1994) found effects of metrical structure. Embedded words preceded by a weak syllable (e.g., *mess* in [nəmɛs]) were detected more easily than embedded words followed by a weak syllable (e.g., *mess* in [mɛstəm]). The MSS predicts such a pattern:

targets in second syllables of weak–strong strings should be easy to detect because the second, strong syllable should be segmented from the first weak syllable by the MSS; in contrast, strong–weak strings with targets in first syllables involve no such segmentation.

A further word-spotting study by Norris et al. (1995) controlled the number of words beginning in the same way as the second syllable, such that some second syllables activated many competitors (e.g., the second syllables of [maskak] and [maskək]), while some activated few competitors (e.g., those of [mintaup] and [mıntəp]). As in the Cutler and Norris (1988) study, the subjects' task was to try to spot words in the first syllable (here, *mask* and *mint*), and as before, they found this harder when the second syllable was strong than when it was weak. Since number of competitors was controlled, this MSS effect cannot be explained as a simple competition effect. The number of second syllable competitors, however, modulated the MSS effect: the difference between strong–strong and strong–weak strings was larger when there were many second syllable competitors.

Vroomen and de Gelder (1995) observed similar effects of number of second syllable competitors in Dutch, in a cross-modal identity priming task. Visual lexical decision to the word *melk* (milk), for example, was faster after the subject had just heard [mɛlkəm] (with no second syllable competitors, and no MSS segmentation at the second weak syllable) than after the subject had heard [mɛlkøm] (with few second syllable competitors, but MSS segmentation at the onset of the second strong syllable), which in turn was faster than after the subject had heard [mɛlka:m] (with many second syllable competitors, and MSS segmentation at the onset of the second strong syllable).

The picture which arises from the current state of research is therefore that the efficiency of human speech recognition relies on multiple solutions to the segmentation problem, involving both exploitation of the information which is available from cues in the signal and in-built characteristics of the process of recognition by competition between simultaneously activated lexical hypotheses. Current models of the recognition process are capable of capturing both these aspects of the system in a unified account.

## MODELING THE RECOGNITION AND SEGMENTATION PROCESS

The current state of the art in theories of word recognition involves computational models within which results from empirical studies can be simulated in detail. A process of competition between simultaneously activated candidate words, as supported by the abundant performance evidence summarized above, is indeed central to models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). Both of these models instantiate the lexical competition process in interactive-activation networks; however they differ in several important ways. For instance, candidate words are selected in Shortlist via an evaluation of the degree to which each word matches *or* mismatches the acoustic-phonetic information in the signal, while TRACE

evaluates words only with respect to their degree of match to the signal, not their degree of mismatch. TRACE thus has no mechanism to account for listeners' sensitivity to mismatching information (Marslen-Wilson & Zwitserlood, 1989). In simulations with Shortlist, the evaluation of match and mismatch is performed via a search of a machine-readable dictionary, allowing the model to draw on a lexicon of over 25,000 words. This large lexicon permits simulations using full sets of experimental materials and realistic estimations of listener vocabularies; TRACE simulations are severely limited in that the available implementation has a lexicon of, at most, only a few hundred words.

In both models, lexical competition operates through inhibitory connections between candidate words. Competition is fiercer between words which overlap with each other more, that is, between words which are fighting over larger sections of the input. After a number of cycles, the activation levels of the candidate words settle into a stable state. If a word has a sufficiently high activation level, it can be recognized. In TRACE, all words in the lexicon are simultaneously considered as candidates; they can in principle all be activated and there are therefore inhibitory connections between all words. Furthermore, the lexicon is reduplicated many times, and each reduplication involves a complete set of words and inhibitory connections. This architecture effectively makes large-lexicon simulations unworkable; of course, it is also highly implausible. In Shortlist, only the words most consistent with the input are considered at any one time. As the model's name suggests, a shortlist of the best candidates is selected; these candidates are wired into a small interactive-activation network, and then compete with each other. Only the inhibitory connections between members of the shortlist are required.

Shortlist deals with the temporal nature of speech through a continuous process of re-evaluation and re-computation (Norris, 1994; Norris et al., 1995). As a simplifying assumption, processing is carried out on a segment-by-segment basis. After a new segment has been presented to the model, the evidence in the signal is re-evaluated, and a new shortlist of candidate words is generated. Depending on degree of bottom-up match and mismatch, words may be added to or deleted from the shortlist. Re-computation of activation levels then proceeds via competition in the interactive-activation network. Old activation levels are not carried forward through re-computations; candidate words compete with one another solely on the basis of their new bottom-up score. The model thus settles on an optimal interpretation of the current input as each segment arrives, without being biased by its previous interpretation. This contrasts with TRACE, where there is no re-evaluation process, and earlier activation levels do influence later ones. This means that in TRACE a candidate word which happens to be highly consistent with an early portion of the input can build up so high an activation level that it is uninfluenced by competition from candidate words aligned with later portions of the input. This does not occur in Shortlist. Thus Shortlist, but not TRACE, can model experimental data showing that the recognition of words is influenced by



competitor words which begin later in time (Norris et al., 1995; Vroomen & de Gelder, 1995).

One further important difference between these competition models is that Shortlist is autonomous (lexical processes cannot influence prelexical processes), while TRACE is interactive (there are top-down connections from lexical to prelexical processing levels). The available evidence on this topic is more consistent with autonomous models (McQueen & Cutler, 1997; Norris, 1994; Pitt & McQueen, submitted). Shortlist thus represents a significant advance over TRACE as a fully implemented competition model. In the present paper we therefore use the Shortlist model as the testbed for simulation of the empirical results. Previous simulation work, summarized below, has shown how the model captures the simultaneous operation of segmentation via competition and metrical effects on segmentation. Given the complexity of the experimental data, it has proved necessary to implement models of spoken word recognition and evaluate their behavior through simulation of that data.

The exploitation of segmentation cues in the input can easily be instantiated in competition models. Norris et al. (1995) implemented the MSS in Shortlist via a combination of two procedures by which strong syllables in the input influenced the activation values of candidate words. The activation of candidates aligned with strong syllables was boosted, and the activation of candidates misaligned with strong syllables was penalized. The combined use of positive and negative information directly mirrored the original formulation of the MSS: the penalty simulated the claim that the input is segmented at the onsets of strong syllables, while the boost simulated the claim that strong syllables are efficient points at which to initiate lexical access attempts. These simulations showed that segmentation and recognition via competition between plausible candidate words is compatible with modulation of the competition process by metrical information in the signal.

Furthermore, the simulations using the Shortlist model with instantiation of the MSS produced the particular prediction that the competition effects should be larger for words embedded in second than in first syllables, and support for this prediction appeared in the results obtained by McQueen et al. (1994). At the [s] of [dʌmɛs], for example, the input is more consistent with *domestic* than with *mess*, simply because the longer word begins earlier, and thus has more segments supporting it. This should produce a strong competition effect, as McQueen et al. found. At the [k] of [sækɹəf], however, there is bottom-up support for both *sack* and *sacrifice*, so there should be weaker competition, and again just this pattern appeared in McQueen et al.'s results.

Thus the current fit between models of the process of word recognition in continuous speech, and experimentally derived evidence for the twin procedures of competition and exploitation of segmentation cues in the input, would seem to have attained a high level. However, as the next section describes, problems still remain to which these procedures cannot offer a solution.

## RULING OUT IMPOSSIBLE SEGMENTATIONS

The above sections have assumed that the listener hears a clearly spoken input and the critical words are part of the listener's vocabulary. In everyday speech, however, either of these conditions may be missing. Often listeners may be presented with input containing words not in their current vocabulary. Errors in production or perception, or the effects of environmental noise, may also lead to inaccuracies in the input representation, with the result that the listener is effectively presented with a word which does not occur in the vocabulary.

Consider what can happen when an error in signal transmission causes a single phoneme to be misperceived; or what can happen when a speaker makes a slip of the tongue and in fact produces a phoneme incorrectly, or for dialect reasons produces a phoneme differently from the listener's previous experience; or even when the next word in the input is one unfamiliar to the listener. Any of these causes—poor quality transmission, imperfect production, dialect mismatch, out-of-vocabulary words—can produce a situation in which part of the input fails to map completely onto a lexical entry. Listeners usually recover from this failure of the mapping process, but the known operations described in the preceding sections do not provide an account of how this occurs. Consequently, the models which incorporate these operations also do not account for successful recognition in such circumstances.

A competition model will in such cases tend to parse the input in terms only of the words that it knows. In some situations this can lead to intuitively unsatisfactory results. Suppose for instance that a speaker with a London dialect produces the phrase *met a fourth time* as *met a fourf time*. Although the word *fourth* will be listed in the model's vocabulary, let us assume that the variant *fourf* is not. And it so happens that if the final segment of *fourf* is ignored, the first three syllables of *met a fourf time* correspond to the word *metaphor*. With such an input there will be considerable bottom-up support for *metaphor*, and a competition model will tend to settle upon the parse *metaphor-?-time*. This parse, however, leaves [f] unaccounted for. Because [f] could not possibly be an English word by itself, this choice implies that the input included a phoneme that was not part of any word.

Note that although this constructed example includes a long word which spans more than one word of the intended input, and although such examples are easy to find (*fill a green bucket* heard as *fill a greem bucket* and leading to the parse *filigree-?-bucket*; *size mix-up* heard as *size mikth-up* giving *seismic-?-up*), the problem is in fact an even more general one. McQueen, Cutler, Briscoe and Norris (1995) show that embedding of words within other words in the English vocabulary is extremely widespread, and that most embedded words overlap with the onset of their matrix words. Thus if (in particular) the final phoneme of a word is misperceived, there is quite a high likelihood that what remains will also be a word—*phoneme* heard as *phonene* gives *phony-?*, *word* perceived as *worb* gives *were/whirr-?*, and so on.

Clearly the tendency of competition models to settle upon parses which leave a phoneme unaccounted for is unattractive in principle and unsatisfactory as a reflection of human word recognition. Human listeners do not generally accept *fourf* as *four* plus *f*. Accordingly, they must be applying some principle or principles which enable them to avoid this situation. Given the great generality of the problem, as a result of widespread word embedding, we find it highly unlikely that all such situations must be resolved by consultation of higher-level contextual information. Instead, we here propose that human listeners can make use of their knowledge of the phonological constraints on the feasibility of stretches of speech as lexical candidates. We suggest that this knowledge is used on-line, during the early stages of word recognition.

Obviously every language in the world has a minimum size for words in the vocabulary. In some languages, the minimal word is relatively large, a fact which has proved important in current developments in phonology (see, e.g., McCarthy & Prince, 1990); for instance Ito (1986, 1990) has argued that the minimal word in Japanese is bimoraic (containing two morae, i.e., consisting minimally of a heavy syllable or two light syllables), while the Australian language Lardil has a bisyllabic minimal word. However, in nearly all languages it is the case that a single consonant cannot form a lexical word. This is certainly true of English; and if listeners were to incorporate this principle in some manner into their analysis of continuous speech input, then they might have at hand a solution to the problems posed by out-of-vocabulary items and corrupted input. That is, they might be able to reject *four* plus *f* as an account of *fourf* not just because *f* is not a member of their vocabulary, but because *f could not ever* be a member of their vocabulary.

Such a constraint—which we can call the Possible-Word Constraint (PWC)—could require simply that wherever possible the input should be segmented so as to produce a string of feasible words. Any parse which results in impossible words (such as isolated consonants) should be rejected in favor of a parse in which all potential lexical candidates are possible words. Thus if a lexical candidate activated by the input begins or ends at a point which leaves a preceding or following stretch of the input consisting solely of a single consonant, the activation of that candidate word should be reduced as a result of this undesirable effect.

A similar proposal has recently been made in a model of initial vocabulary acquisition by infants. Consider that prelinguistic infants are effectively listeners who encounter “out-of-vocabulary” items rather more often than most listeners do. In the process of acquiring a vocabulary, infants would be well advised to avoid any temptation to consider consonants, or consonant clusters, as potential words to be added to the lexicon. Brent and Cartwright (1996) conducted a series of computational studies of vocabulary acquisition, in which they used a Minimum Representation Length (Rissanen, 1989) technique to learn a vocabulary from selections of the CHILDES database (MacWhinney & Snow, 1985), transcribed in such a way that all word boundary

markings were deleted. Brent and Cartwright showed that vocabulary acquisition was improved by adopting a strategy which insisted that all words contained a vowel, thus effectively ruling out consonants as feasible candidate words. The benefits of this “vowel constraint” were in addition to benefits gained from observing phonotactic constraints on permissible clusters appearing at word boundaries.

In the experiments that follow, we explicitly test whether human listeners command a constraint on what could be a possible word. Indirect evidence suggesting that such a constraint may play a role in human word recognition arose from observations that we made during previous work. The word-spotting study of Cutler and Norris (1988) included the item [ɔʒʌmpəv]. All subjects who responded to this item reported hearing the word *jump*, exactly as we had intended. However, in our Shortlist simulations of that study we observed that the word most strongly activated by [ɔʒʌmpəv] was *juniper* ([ɔʒʌmpə] in British English). When we constructed the materials for that experiment we failed to notice that *juniper* was embedded in [ɔʒʌmpəv]; and our subjects likewise failed to notice it. Shortlist simply recognized the longest word consistent with the input; human subjects (and experimenters) recognized the shorter word, leaving the syllable [əv] unaccounted for, rather than the longer word, which would have left the isolated phoneme [v]. It is of course possible that this listener bias reflected some factor such as word frequency, or the specific task demands of the experiment. But it is also interpretable as a preference not to postulate word boundaries at points which result in impossible residue words—that is, as a PWC.

In the experiments reported below, we chose not to undertake a more systematic investigation of which response is elicited by strings like [ɔʒʌmpəv] containing two embedded words, since it is impossible to equate two such potential response items on all relevant variables. Instead, listeners performed the word-spotting task on strings containing a single embedded word, and we measured the relative speed and accuracy of response to the word as a function of whether what was left over in the input was a single consonant, that is, an impossible candidate for word status, or a CVC sequence, that is, a potential candidate word. Thus listeners’ response time and accuracy in spotting the embedded word *sea* was compared in *seash* [siʃ] versus *seashub* [siʃʌb]; spotting of *apple* was compared in *fapple* [fæpəl] versus *vuffapple* [vʌfæpəl]. If listeners indeed have a bias against placing word boundaries at points that result in impossible-word residues, it should be far harder to spot these words in [siʃ] and [fæpəl] than in [siʃʌb] and [vʌfæpəl].

## EXPERIMENT 1

### *Method*

*Subjects.* Subjects were 42 students from Downing College, Cambridge who were each paid 3 pounds for participating.

*Materials.* Stimuli were based on a set of 96 words, half of which were monosyllabic and half of which were bisyllabic. All bisyllabic words had a strong-weak stress pattern. Every word was effectively embedded into two nonsense strings by the addition of some phonetic context; the most important manipulation in the experiment was that in one nonsense string for each word this context formed a possible, though nonexistent word of English (always a CVC, as in *vuffapple* or *seashub*), while in the other it was impossible as a word (always a single consonant, as in *fapple* or *seash*). Forty-eight of the items (24 monosyllabic, 24 bisyllabic) were preceded by their context (*maffegg*, *fegg*, *vuffapple*, *fapple*); the other 48 were followed by their context (*seashub*, *seash*, *sugarthim*, *sugarth*).

The nonsense strings contained no embedded words other than the target item itself (with the necessary exception of single phoneme words such as *a*). All added possible-word contexts were strong syllables, containing the lax vowels [æ], [ɛ], [ɪ], [ɒ], [ʌ] (as in British English *hat*, *het*, *hit*, *hot*, *hut*). All diphone transitions used in the contexts were present in the English vocabulary (Gimson, 1980). To minimize allophonic variation of the target words in their two different contexts, all words with Preceding Context began with vowels, and all words with Following Context ended with [l], [n] or a vowel (note that *sugar*, for example, ends with a vowel in standard Southern British English, which is a non-rhotic dialect). Also, in order to facilitate cross-splicing of targets in Experiment 2, we avoided combinations of word and context with transitions difficult to locate in the waveform (e.g., [lm]). The materials are listed in full in the Appendix.

In addition to the 96 experimental items there were 192 filler items containing no embedded words. Fillers were constructed along similar lines to the experimental items. In particular, filler strings contained the same proportion of one-, two-, and three-syllable items as the experimental items. Half of the fillers were in fact generated by starting from a monosyllabic or bisyllabic nonword and then adding a consonant or CVC in exactly the same way as for the target items.

Materials were read from a broad phonetic transcription of all items by a male speaker of standard Southern British English and recorded onto Digital Audio Tape, sampling at 44.1 kHz. They were then digitally downsampled to 22.05 kHz with 16 bit resolution and examined with a speech editor. Timing pulses were aligned with the onset of each target word. Final experimental tapes were produced by upsampling the signal to 44.1 kHz. The speech was recorded on the left channel of a Digital Audio Tape and timing pulses (inaudible to the subject) were recorded on the right channel.

Four experimental lists were constructed, two consisting only of items with Preceding Context and the other two only of items with Following Context. Each of these pairs of lists contained 12 items in each of the Word Length (Monosyllabic, Bisyllabic) by Context Type (Possible, Impossible) conditions; Context Type was counterbalanced across each pair of lists. All subjects heard one of the Preceding Context lists and one of the Following Context lists; one quarter of the subjects heard each of the possible combinations of one Preceding Context and one Following Context list. Order of presentation of Preceding and Following Context was also counterbalanced, making eight subject groups in all.

Each half of the experiment was preceded by 12 practice items, including four target items with context in the appropriate position (one similar to each Word Length by Context Type condition) and eight filler items.

*Procedure.* Subjects were tested individually in a quiet room. They were given written instructions explaining that they would hear a list of nonsense words and that they should press the button in front of them as quickly as possible whenever they heard a nonsense word beginning (or ending) with a real word. After pressing the button subjects had to say the word they had spotted out aloud. After the practice block subjects were given further written instructions indicating which items they should have responded to. Subjects' vocal responses were recorded on cassette tape for further analysis. There was a pause half way through the experiment at which point subjects were given instructions and practice for the second half of the experiment, that is, were instructed that they should now listen for words embedded at the end (or beginning) of the nonsense words. Stimuli were presented over Sony CD550 headphones and subjects' responses were logged by a Commodore microcomputer.

TABLE 1  
 Mean Reaction Time (RT, in Milliseconds), Measured from Target-Word Offset,  
 and Mean Percentage Error Rates (Err), in Experiment 1

Context type	Target word length							
	Monosyllabic				Bisyllabic			
	Preceding context		Following context		Preceding context		Following context	
	RT	Err	RT	Err	RT	Err	RT	Err
Possible	920	57%	870	28%	566	14%	829	17%
Impossible	1026	52%	965	39%	650	18%	813	38%

### Results and Discussion

Raw Reaction Times (RTs), measured from target-word onsets, were adjusted by subtracting word lengths to yield RTs from word offset. The tape recordings of subjects' spoken responses were analyzed to identify occasions when subjects pressed the button but then either failed to make a verbal response or responded with a word other than the appropriate target. These responses, along with responses shorter than 200 ms or greater than 1750 ms, as measured from target offset, were treated as errors. In some previous word-spotting experiments we have rejected any subjects failing to detect a specific percentage of targets (e.g., 50% in McQueen et al., 1994, Experiment 1). Because of the high error rates observed in the present study (mainly in the Impossible Context condition) we rejected only subjects who failed to identify any targets in one or more conditions. Two subjects were discarded in this way, leaving 40 subjects, of which 20 had heard each list. We also found that no subject responded to the targets *eve* (Monosyllable, Preceding Context) in the Possible Context condition, or *pole* (Monosyllable, Following Context) in the Impossible Context condition, so these two words were not included in any of the analyses. The mean RTs and error rates are shown in Table 1.

Analyses of variance were performed on RTs and errors with both subjects ( $F_1$ ) and items ( $F_2$ ) as the repeated measure. We report only effects significant in both analyses. For the RTs, the main effect of Context Type was highly significant: words with Possible Context were responded to 67 ms faster than words with Impossible Context ( $F_1(1,32) = 19.85, p < .001$ ;  $F_2(1,90) = 13.63, p < .001$ ). There was also a significant effect of Word Length, with Bisyllabic words being responded to 231 ms faster than Monosyllabic words ( $F_1(1,32) = 302.37, p < .001$ ;  $F_2(1,90) = 62.06, p < .001$ ), and a significant effect of Context Position, with Following Context slowing responses more than Preceding Context ( $F_1(1,32) = 14.51, p < .001$ ;  $F_2(1,90) = 4.43, p < .05$ ). However, this latter effect was qualified by a two-way interaction between Context Position and Word Length ( $F_1(1,32) = 105.56, p < .001$ ;

$F_2(1,90) = 12.84, p < .001$ ); for Bisyllabic words Preceding Context resulted in faster responses than did Following Context, but the opposite was true for Monosyllabic words.

Subjects' errors showed the same two principal main effects which had appeared in their RTs. Responses to words with Possible Context were 8% more accurate than responses to words with Impossible Context ( $F_1(1,32) = 13.52, p < .001$ ;  $F_2(1,90) = 8.29, p < .01$ ). There was a significant effect of Word Length, responses to Bisyllabic words being 21% more accurate than responses to Monosyllabic words ( $F_1(1,32) = 141.89, p < .001$ ;  $F_2(1,90) = 40.90, p < .001$ ). Words with Following Context were responded to 4.7% more accurately than words with Preceding Context, but this effect did not reach significance in both analyses. Once again there was a two-way interaction between Context Position and Word Length ( $F_1 = 95.91, p < .001$ ;  $F_2(1,90) = 21.25, p < .001$ ); for Bisyllabic words, Preceding Context led to more accurate responding than Following Context did, but the opposite was true for Monosyllabic words.

As predicted, responses in this experiment to targets with Possible Context were both faster and more accurate than responses to targets with Impossible Context. However, there remains a possibility that this effect is attributable to differences in the acoustic-phonetic realization of the targets in the different contexts rather than to the direct effects of the context itself. Perhaps targets with Possible contexts are closer to the canonical form of the target word than are targets with Impossible contexts. To assess this possibility, Experiment 2 repeated Experiment 1 with the target words cross-spliced between the Possible and Impossible Context conditions. For example, *sea* from *seashub* was exchanged with *sea* from *seash*, and vice versa. If the advantages of CVC context in Experiment 1 were due to differences in the realization of the targets, then the Context Type effect should be reversed in Experiment 2. If the Context Type effect is instead attributable to the Possible-Word Constraint, the results of Experiment 2 should be the same as Experiment 1.

## EXPERIMENT 2

### *Method*

*Subjects.* Subjects were 43 students from St. John's College, Cambridge, who were each paid for participating in the study.

*Stimuli.* The materials for Experiment 2 were identical to those from Experiment 1 but with the target words cross-spliced between Possible and Impossible Context conditions. All cross-splicing was performed at zero-crossings. Because of the care taken in the initial construction of the materials to minimize allophonic variation, the spliced items sounded natural and the splices appeared to provide no cues to distinguish between fillers and targets.

*Procedure.* The procedure was identical to Experiment 1.

### *Results and Discussion*

As in Experiment 1, subjects who made no correct responses in one or more conditions were rejected. Three subjects were rejected on this criterion,

TABLE 2  
 Mean Reaction Time (RT, in Milliseconds), Measured from Target-Word Offset,  
 and Mean Percentage Error Rates (Err), in Experiment 2

Context type	Target word length							
	Monosyllabic				Bisyllabic			
	Preceding context		Following context		Preceding context		Following context	
	RT	Err	RT	Err	RT	Err	RT	Err
Possible	868	58%	867	33%	545	12%	786	19%
Impossible	1098	63%	938	40%	626	23%	842	54%

leaving 20 in each list condition. Three words were removed from the analysis because they were not detected by any subject in one condition. The words were *eve* and *pole*, which were also removed from the analysis of Experiment 1, and *ebb*. Although in this experiment all three of these words were never detected in the Possible Context condition, removal of these words did not alter the pattern of significance in the analysis. RTs were again adjusted so as to measure from target-word offset, and responses outside the 200–1750 ms window were treated as errors. The mean RTs and error rates are shown in Table 2.

Words with Possible Context were responded to 109 ms faster than words with Impossible Context ( $F_1(1,32) = 85.10, p < .001$ ;  $F_2(1,89) = 25.07, p < .001$ ). Bisyllabic words were identified faster than Monosyllabic words ( $F_1(1,32) = 355.21, p < .001$ ;  $F_2(1,89) = 55.95, p < .001$ ). The effect of Context Type was greater for words with Preceding Context than for words with Following Context ( $F_1(1,32) = 9.21, p < .005$ ;  $F_2(1,89) = 13.79, p < .001$ ), and for Monosyllabic than for Bisyllabic words ( $F_1(1,32) = 12.07, p < .005$ ;  $F_2(1,89) = 4.23, p < .05$ ). There was also a two-way interaction between Word Length and Context Position ( $F_1(1,32) = 121.12, p < .001$ ;  $F_2(1,89) = 13.63, p < .001$ ): for Bisyllabic words Preceding Context resulted in faster responses than Following Context, but the opposite was true for Monosyllabic words.

As in Experiment 1, subjects' errors showed the same main effects as their RTs. Responses to words with Possible Context were 17% more accurate than responses to words with Impossible Context ( $F_1(1,32) = 138.43, p < .001$ ;  $F_2(1,89) = 50.38, p < .001$ ). Responses to Bisyllabic words were 19% more accurate than Monosyllabic words ( $F_1(1,32) = 122.31, p < .001$ ;  $F_2(1,89) = 26.51, p < .001$ ). The effect of Context Type was greater for words with Preceding Context than for words with Following Context ( $F_1(1,32) = 31.01, p < .001$ ;  $F_2(1,89) = 14.23, p < .001$ ). The Context Type effect was larger for Bisyllabic than for Monosyllabic words ( $F_1(1,32) = 21.57, p < .001$ ).



.001;  $F_2(1,89) = 6.70, p < .05$ ). Once again there was an interaction between Word Length and Context Position: for Bisyllabic words Preceding Context led to more accurate responding than did Following Context, but the opposite was true for Monosyllabic words ( $F_1 = 273.50, p < .001$ ;  $F_2(1,89) = 30.62, p < .001$ ).

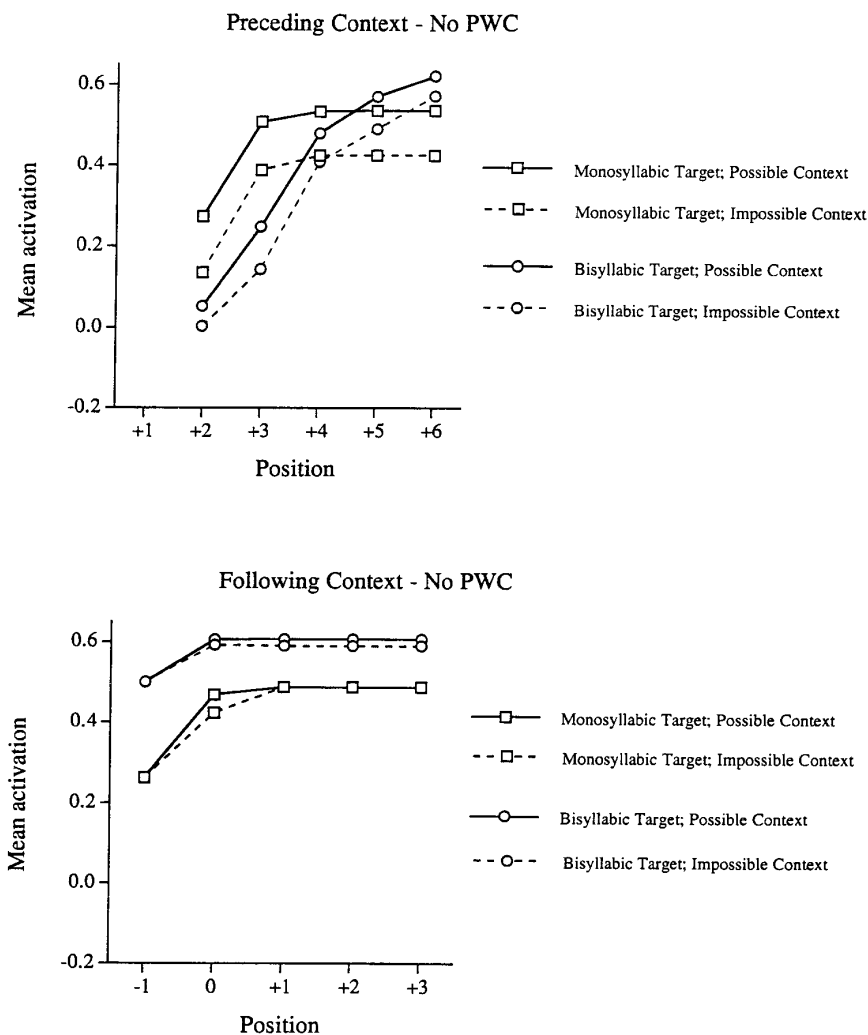
Thus the results of the present experiment confirm that listeners consistently find it easier to spot words embedded in Possible than in Impossible contexts. Experiment 2 has shown that the advantage of Possible Context over Impossible Context which we observed in Experiment 1 was not due to any differences in the acoustic realization of the words in the two conditions. The Possible Context advantage remained even when the target words were cross-spliced between CVC and consonantal contexts. Together the experiments provide powerful evidence for the claim that listeners do not entertain segmentations which would lead to impossible words, that is, they lend support to the proposal that a Possible-Word Constraint plays a role in listeners' segmentation performance. Thus it is not the case that consultation of higher-level context is the only option via which listeners can rule out impossible segmentations: these can be rejected immediately if they involve phonological forms which could not possibly be words.

## MODELING THE PWC

### 1. The Possible-Word Constraint

Given the convincing evidence in support of a Possible-Word Constraint (PWC) provided by the present experimental results, we turn our attention next to the question of how to integrate these findings in a computational model of spoken-word recognition. As we described in the introduction, we will use the Shortlist model (Norris, 1994) as the testbed for answering this question, because it is available in a form in which all of our experimental materials can be incorporated in the simulations, and the entire English vocabulary can participate in the provision of word candidates. Furthermore, it is already known that the model accurately captures the effects of competition between activated candidate words in continuous speech input, including the effects of varying numbers of competitors, and also accurately models the effects of metrical structure of the input observed in earlier experiments. Note that the effects of competition, and the interaction of competition effects with metrical and other known effects, cannot be simply predicted, because they depend, for each member of the set of items used in an experiment, upon that item's precise competitor environment—all the words in the English language which overlap with any portion of the item. Only realistic simulations with a large lexicon allow us to assess how accurately the model simulates the way we know listeners perform.

Figure 1 shows simple simulations of the present data with the Shortlist model as described in Norris et al. (1995), using a lexicon of 26455 words, based on the *Longman Dictionary of Contemporary English* (Procter, 1975).



**FIG. 1.** Mean target activation levels for the materials from Experiments 1 and 2 in Shortlist. The model was run on a 26455-word English lexicon, using the implementation of the Metrical Segmentation Strategy (MSS) described in Norris et al. (1995). The mean activation functions are shown for the monosyllabic and bisyllabic target words, with both possible and impossible contexts. The upper panel has plotted the data for targets with preceding context (such as *apple* in *fapple* and *vuffapple*). The lower panel has plotted the data for targets with following context (such as *sugar* in *sugarth* and *sugarthim*). The activation functions are aligned relative to the last/first phoneme of the context (0). Thus, for targets with preceding context, +2 is the second segment of the target word, while for targets with following context, +1 is either the first segment of the possible context or the first silent segment following the impossible context.

All subsequent simulations, unless otherwise stated, also make use of this large lexicon. It can be seen that for words with Following Context this 1995 version of Shortlist produces equal activation in Possible and Impossible Contexts, and thus implicitly predicts that word-spotting should be equally difficult irrespective of whether the context is Possible or Impossible. Such a prediction is clearly in conflict with the results which we have observed. However, in agreement with the human data, the simulation correctly predicts that when the word has Preceding Context, the Possible Context should be easier than the Impossible Context.

The difference between the Preceding Possible and Impossible Contexts in this simulation follows from the operation of the Metrical Segmentation Strategy within Shortlist (Norris et al., 1995). The MSS gives a boost to words with strong onsets that begin where there is the onset of a strong syllable in the input. In the Possible Contexts these items receive a boost in the normal fashion. However, in Preceding Impossible Contexts the embedded words no longer begin at strong syllable onsets and therefore receive no boost. Words with Preceding Impossible Contexts achieve a lower level of activation (and should therefore be harder to detect) than target words with Preceding Possible Contexts. There is no difference between the Possible and Impossible Following Contexts, on the other hand, because the MSS simply has nothing to say about word offsets.

Thus although the MSS boost does lead to a difference between Possible and Impossible Contexts being predicted in the case of Preceding Context, this mechanism is not able to provide a complete account of the effects we have observed, in that it leads to the incorrect prediction that detection of targets with Following Impossible Contexts should be little harder than detection of the same words with Following Possible Contexts. In fact, both types of words are effectively predicted to be little harder to detect in these Following Contexts of either type than simply in isolation. Words in isolation also do not receive a boost from the MSS. The experiment, though, showed that words in Impossible contexts are very difficult indeed to detect irrespective of the position of the context; the word-spotting results show long RTs and extremely high error rates for Impossible Context in both Preceding Context and Following Context conditions.

That is, the experimental effects we have observed show a fundamental difference between candidate segmentations as a function of whether a CVC or a consonantal portion of the input is involved; they suggest that segmentations which allow nonviable residues are intrinsically disfavored. There is no difference, however, in this effect as a function of the context position. The results thus seem to demand an explanation whereby candidate words in impossible contexts, wherever they occur, are by their very occurrence penalized.

How can a PWC of this type be modeled computationally? One potential problem is that the PWC can be seen as applying to the entire parse of the input rather than to individual lexical candidates. Superficially, the PWC might appear to be most readily incorporated into a speech recognition system

that operated by computing alternative paths (word strings) through a word lattice in the manner of many automatic speech recognition systems (Chien, Lee, & Chen, 1991; Thompson, 1990). The PWC could then be applied by reducing the evidence in favor of any path that violated the constraint. However, the basic unit of computation in a competition model such as Shortlist is the word rather than the path. Although the outcome of the competition process is a single interpretation (path through the input), the model does not undertake direct comparisons of the likelihood of alternative paths; all computations concern individual candidate words.

In order to incorporate the Possible-Word Constraint into Shortlist in the form of disfavoring of parses with unacceptable residues, it is necessary to be able to make individual candidate words sensitive to whether or not their boundaries are going to lead to a violation of the constraint. A straightforward way of achieving this is to penalize any candidate word that leads to a word boundary being placed such that there is no vocalic segment between that boundary and the next known boundary. In this form, we implemented the PWC in the Shortlist model.

Shortlist does not itself perform a syllabic parsing of its input. The input to Shortlist consists of a phonemic transcription, with strong syllable onsets explicitly marked, as in the Norris et al. (1995) simulations. The PWC implementation therefore consists in allowing Shortlist to determine, as each successive input phoneme is presented to the program, whether that phoneme is a possible word boundary according to the procedure described above. That is, if there are only consonants between the segment at the boundary of the current candidate word and the next boundary marked in the input (working both forwards and backwards) then the current segment is *not* at a possible boundary.

Any candidate with a boundary at a point which violates the PWC then has the bottom-up evidence in its favor reduced. Thus in [siʃ] the candidate word *sea* has its activation reduced because only the consonant [ʃ] occurs between its boundary and the silence which itself is a clear boundary signal. Note that thus *sea* only violates the PWC when the silence following the third segment in [siʃ] is presented. Until that point, the input could continue as a syllable (as in [siʃʌb] in which *sea* does not violate the PWC). The PWC comes into effect as the next boundary is computed, silence here constituting a clear and unambiguous cue to a syllable boundary.

This simple method of implementing the PWC simulates the results of Experiments 1 and 2, as Figure 2 shows. Activation levels are considerably lower for target words with Impossible Contexts than for target words with Possible Contexts, both for Preceding and Following Contexts and for both monosyllabic and bisyllabic words. Words in Impossible Contexts are penalized because their boundaries violate the Possible-Word Constraint, whereas words in the Possible Contexts are not penalized. As one would expect, the difference between Possible and Impossible Following Contexts emerges only after the first segment following the end of the word, as this is the first point

where Shortlist can determine that the Following Impossible Context violates the PWC. Note that the simulation in Fig. 2 uses only the PWC as described here, and not the instantiation of the MSS as described by Norris et al. (1995).

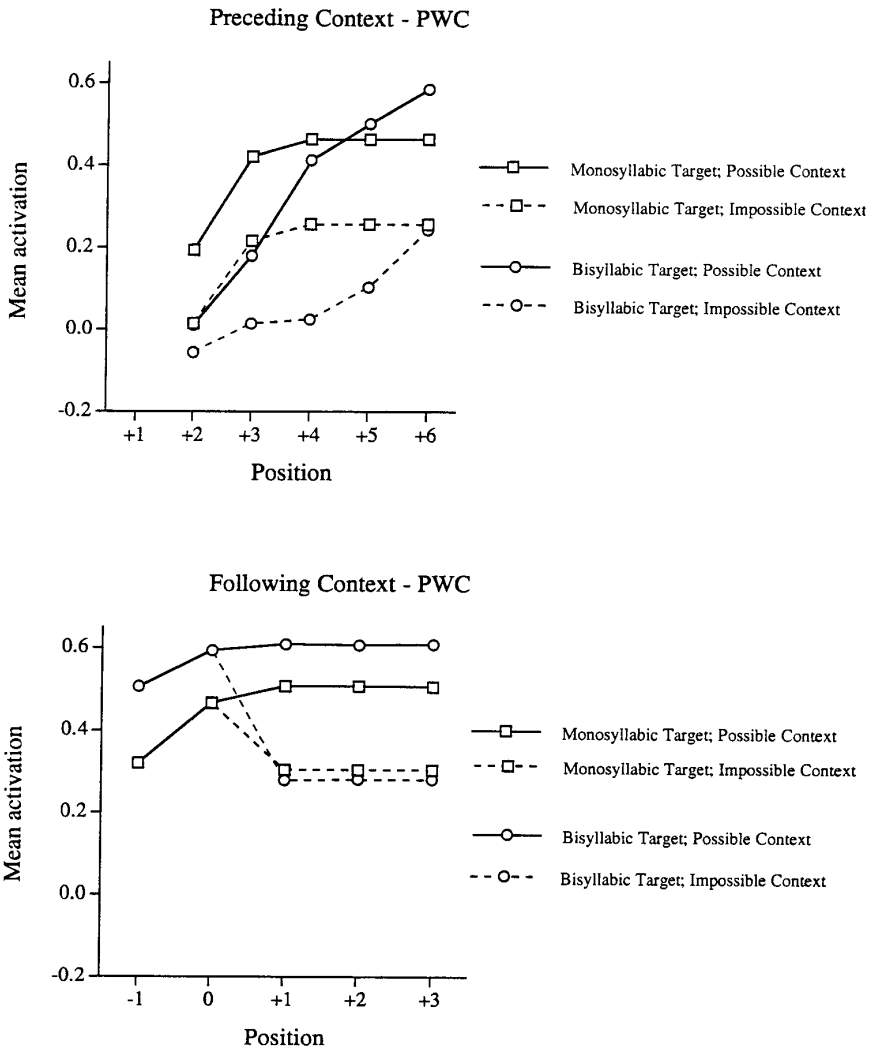
## 2. Imperfectly Specified Input

The PWC is also effective in enabling the recognition system to deal with poorly-specified or out-of-vocabulary input. As we described in the introduction, continuous speech recognition requires a mechanism to deal with speech input which, for a number reasons, does not map neatly onto lexical entries. A speaker may use words (or dialectal variants of words) which are not in the listener's vocabulary. Background noise, slips of the tongue, or other factors could produce speech as input for comprehension that is poorly specified, and hence also mismatches with lexical entries. For example, the phrase *met a fourth time* could be heard as *met a fourf time* either because of dialectal use of the pronunciation *fourf* or because the final fricative ([θ]) was distorted by background noise, leading to perception of [f]. Whatever the cause of imperfectly-specified input, the recognition system would be well-served by a mechanism which ruled out spurious, and impossible, alternative parses of this type of input (such as *metaphor f time*). The PWC provides exactly such a mechanism.

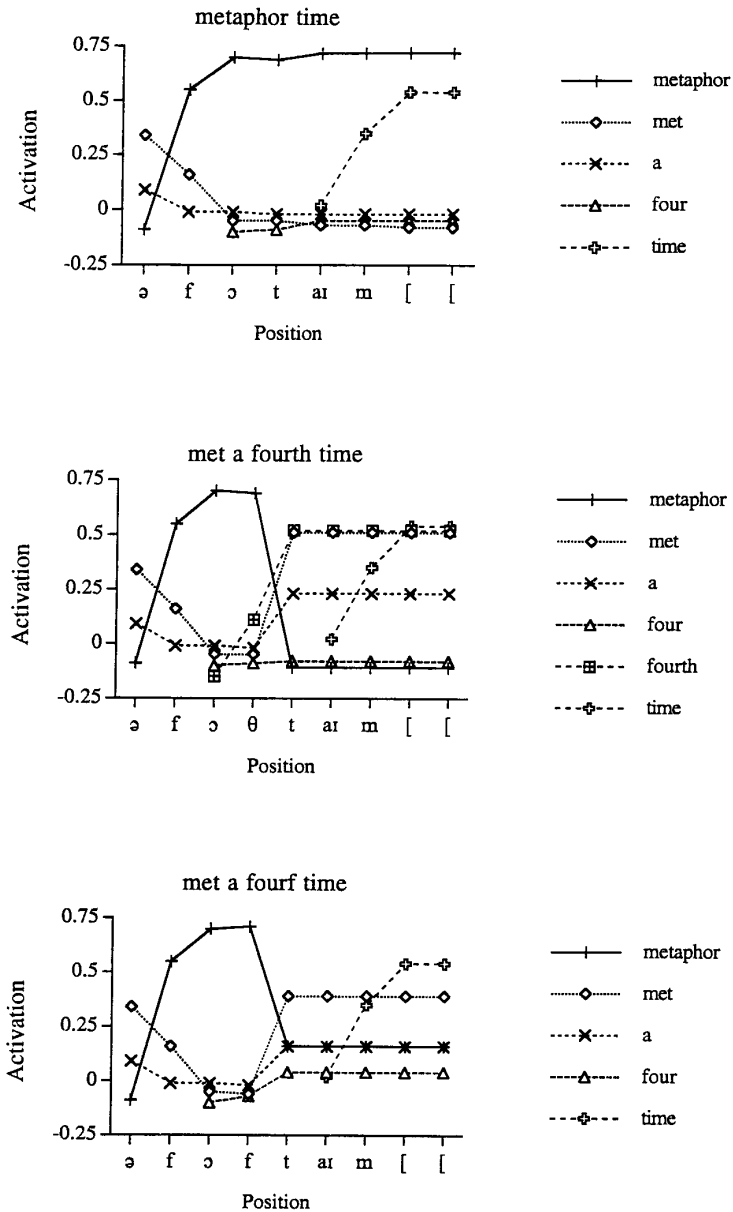
Figure 3 shows the results of simulations given three different inputs (*metaphor time*, *met a fourth time*, and *met a fourf time*), with the PWC in Shortlist, here using, for greater phonetic detail, a lexicon of 8932 words based on the CELEX English lexicon (Burnage, 1990). Given the input *metaphor time* (upper panel), the candidate words *metaphor* and *time* dominate the activation pattern. Although words like *met*, *a* and *four* are considered, they are rejected in favor of *metaphor*. With the current activation parameters of the model, Shortlist has the property of favoring one longer word over a string of shorter words embedded within that longer word.

The PWC has little role to play given an input like *metaphor time*. But given the input *met a fourth time*, the PWC acts to disfavor the candidate *metaphor* (see the central panel of Fig. 3). Here it exploits the fact that the word *time* consists of a strong syllable. Since strong syllables are likely word onsets, a likely boundary is signalled between the [θ] and the [t]. For the PWC, the candidate *metaphor* is then misaligned with this boundary because there is no vowel between the offset of *metaphor* and the boundary (there is only the [θ]). So the longer word, that would be consistent with a parse containing an impossible word (*metaphor th time*), is disfavored, and the candidates *met*, *a*, *fourth*, and *time* can dominate the activation pattern.

Finally, consider the imperfectly-specified input *met a fourf time* (lower panel of Fig. 3). The PWC can again be seen to have the desired effect of preventing the spurious recognition of *metaphor*. As in the *met a fourth time* example, *metaphor* is misaligned with the boundary cued by the strong syllable of *time*, now leaving the impossible word *f*. But note that the activation of *metaphor* remains at a higher level than in the previous example. This is



**FIG. 2.** Mean target activation levels for the materials from Experiments 1 and 2 in Shortlist. The model was run on a 26455-word English lexicon, using the Possible-Word Constraint (PWC). The mean activation functions are shown for the monosyllabic and bisyllabic target words, with both possible and impossible contexts. The upper panel has plotted the data for targets with preceding context (such as *apple* in *fapple* and *vuffapple*). The lower panel has plotted the data for targets with following context (such as *sugar* in *sugarth* and *sugarthim*). The activation functions are aligned relative to the last/first phoneme of the context (0). Thus, for targets with preceding context, +2 is the second segment of the target word, while for targets with following context, +1 is either the first segment of the possible context or the first silent segment following the impossible context.



**FIG. 3.** Activation levels of words given the phrases *metaphor time* (top panel), *met a fourth time* (middle panel), and *met a fourf time* (bottom panel) in Shortlist. The model was run on an 8932-word English lexicon, using the Possible-Word Constraint. Activation values are given for *metaphor*, *met*, *a*, *four*, and *time* in all three panels, and for *fourth* in the middle panel (this word failed to become significantly activated in the other two simulations). The activation functions are plotted for the segments of each phrase, beginning at the schwa ([ə]) of *metaphor* and *a*, and for two following silent segments (“[’]”).

because there is no complete alternative parse (like *met a fourth time*) which can dominate the activation pattern (because *fourf* is not in the Shortlist lexicon). Both *met* and *a* remain as highly activated as they were in *met a fourth time*, as does *time*, leaving the syllable *fourf* for which there is no dominant lexical hypothesis (*four*, like *metaphor*, is misaligned with the boundary before *time* and is therefore also penalized by the PWC). In other words, the PWC causes Shortlist to rule out alternatives which leave impossible words (like *f*) in the parse of the input, and to prefer other alternatives where they are available (like *met*, *a* and *time*), leaving sections of the input (that are at least as long as one syllable) which are possible (albeit currently unrecognizable) words.

Thus the PWC offers a very realistic approach to solving the problem of imperfect input and out-of-vocabulary words. Precisely the correct segmentation of *met a fourth time* is achieved, even though no lexical candidate can be assigned to the mispronounced (or misperceived) string *fourf*. Note that the choice of the best candidate given *fourf* can only be accomplished via extra knowledge, which Shortlist at the moment cannot call upon. For instance, Shortlist currently incorporates no way of exploiting the acoustic similarity of [f] and [θ], which might well be considered potentially relevant by a human listener. Nor can Shortlist (or any other model of spoken-word recognition that we know of) yet add a new entry to the lexicon, as a human listener might choose to do (for instance upon hearing *I met a forf today*). Nevertheless the instantiation of the PWC as an on-line constraint on the activation of candidate words accurately captures the subjective and experimental human data—for listeners, the parse *four + f* simply never counts as a viable recognition option.

The operation of the PWC as we have implemented it exploits to maximum effect the clear and unambiguous information about syllable boundaries provided by the silence at either end of a stretch of speech input. But as we saw with the *met a fourth time* simulations in this section, the PWC is also able to exploit the marking of strong syllables in the Shortlist input in exactly the same way as it exploits silence; it treats a strong-syllable onset as a boundary and evaluates candidate words with respect to the material intervening between their edges and such a known boundary. In other words, the concept of a known boundary for the purposes of the PWC is generalisable; not only the unambiguous silence cue can be used, but also the more probabilistic information provided by metrical structure.

This raises the issue of the extent to which the operation of the PWC might also be generalisable to other operations than those for which it was explicitly designed, which were to account for the present experimental results and for the known behaviour of human listeners with imperfectly specified input. For instance, we saw in the first set of simulations above an indication of some overlap between the consequences of the PWC and the consequences of the independently motivated simulation of the Metrical Segmentation Strategy by Norris et al. (1995). Thus the PWC clearly penalizes *apple* in [fæpəl]. But



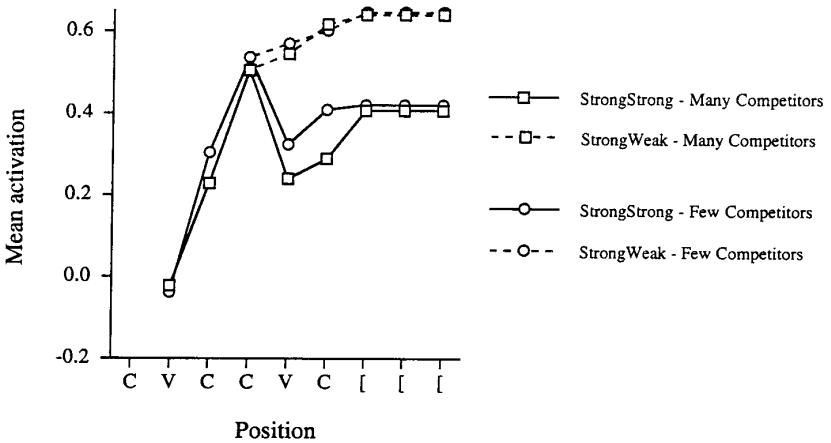
the 1995 MSS implementation also penalized a word with a strong onset that was not aligned with a strong onset in the input—as is the case with *apple* in [fæpəl]. Here the PWC would seem to be doing (with an essentially simpler mechanism) some of the work of the MSS. Norris et al. (1995) were forced to implement the MSS with a two-part mechanism of boost for strong-initial candidate words plus penalty for misaligned words. Could inclusion of the PWC perhaps enable a simpler simulation of the data supporting the MSS?

Note that Norris et al. (1995) chose the two-part implementation of the MSS after exploration of each part separately; they established that simply applying a penalty to candidate words misaligned with strong syllables in the input did *not* by itself succeed in capturing their experimental results plus the data from McQueen et al. (1994). However, the penalty applied by the PWC is more general than that incorporated in the MSS instantiation, first in that it is triggered by any boundary and not just a strong syllable, and second in that it strikes not only the misaligned candidate word but any other candidate word resulting from the same parse. In the next section we investigate the performance of the PWC in Shortlist with the materials which, in a number of experiments, have produced clear effects of metrical structure. Again, simulation with the whole English vocabulary is necessary to ascertain precisely how the metrically varying materials will behave under competition.

### 3. The Possible-Word Constraint and the MSS

#### a. Norris, McQueen, and Cutler (1995)

This experiment, plus that of Vroomen and de Gelder (1995) in Dutch, without doubt provide the most informative data on metrical effects and their relation to competition processes. Both of these studies showed that metrical effects (e.g., slower recognition of *mint* in strong–strong (SS) strings like [mɪntaup] than in strong–weak (SW) strings like [mɪntəp]) are larger when there are more competitors beginning at the final consonant of the target word. The metrical effect itself comes about because the second syllable indicates that there should be a syllable boundary before the [t] of *mint* in the SS string (in which the second syllable contains a strong vowel) but not in the SW string (with a weak vowel in the second syllable). According to the MSS, candidates ending on the [t] in the SS string are disfavored because they straddle a syllable boundary which is likely to be a word boundary. In the 1995 implementation of the MSS, such candidates are penalized because the [t] begins a strong syllable but the [t] in *mint* is not marked in the lexicon as being the onset of a strong syllable. Additionally, competitors beginning with the [t] are boosted because they begin at a strong syllable, and these competitors cause more inhibition of *mint* than competitors in SW strings that are not boosted. Given the operation of lexical competition in Shortlist, the size of the difference between activation levels of target words in SS and SW strings is modulated by the number of second syllable competitors. As we saw, this modulation of the metrical effect by the number of second syllable competitors is exactly what was observed in the human data.



**FIG. 4.** Mean target activation levels for the materials from Norris et al. (1995). The model was run on a 26455-word English lexicon, using the Possible-Word Constraint. The mean activation functions are shown for target words in StrongStrong strings with many second syllable competitors (e.g., *mask* in [maskak]), in StrongWeak strings also with many second syllable competitors (e.g., *mask* in [maskək]), and in both StrongStrong and StrongWeak strings with few second syllable competitors (e.g., *mint* in [mɪntaʊp] and [mɪntəp], respectively). The activation functions are aligned with the Consonant/Vowel (C/V) structure of the target-bearing strings and three following silent segments ('['], such that, for example, the [k] of *mask* or the [t] of *mint* are aligned with the second C of the medial consonant cluster.

Assuming that strong syllable onsets are marked in the input and that they are considered to be highly probable word boundaries, then the PWC produces an effect very similar to the (inhibitory component of the) MSS for words such as *mint* in the SS string [mɪntaʊp]. Here, *mint* will be penalized by the PWC because it ends at a point that cannot be a word boundary (there is a known boundary before the [t], at the onset of the strong syllable, but no vowel between this boundary and the end of *mint*; there is only the [t]). Norris et al. (1995: 1219) reported Shortlist simulations of their data using only the penalty component of the MSS, and these simulations did appear to capture the observed data pattern well, although the analysis of the simulations revealed that the interaction between stress and number of competitors was not quite significant. Figure 4 shows Shortlist simulations of the same Norris et al. data using only the PWC, without any component of the MSS.

It can be seen that the PWC alone, without any further MSS instantiation, in fact succeeds in capturing the interaction between the metrical effect and the number of competitors. The difference in activation levels between SS and SW strings is greater when there are more competitors. Moreover, the interaction between stress pattern and number of competitors is significant (at the final consonant:  $F(1,38) = 8.93, p < .005$ ). In Norris et al.'s (1995) simulations this was the case only with the combined boost and penalty version of the MSS. However the PWC (although it operates via a penalising

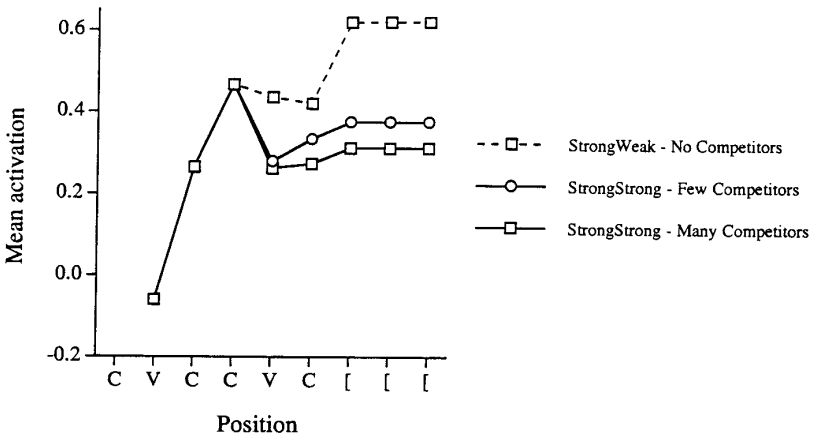
mechanism) is far stronger than the 1995 MSS penalty was, and in effect even operates to produce effects similar to those of the 1995 penalty and boost combined. The 1995 MSS penalised only words misaligned with a strong onset. The PWC penalises all words which in any way leave an impossible word; the effect of this is that not only are words misaligned with strong onsets penalised, but also competitors for part of their input are rendered stronger by having, in turn, their competitors penalised. The way this works is as follows. The candidate word *mint* is penalized because it ends one consonant after a signalled boundary, and thus appears to consist of two portions, one well-formed ([mɪn]) and one ill-formed ([t]). However, the PWC also acts to penalize word candidates beginning at the same point one consonant after a signalled boundary, such as, in this case, all words beginning [au]; they would appear to leave the impossible unparsed residue [t]. Any candidate words beginning with the [t] therefore undergo less competition from words beginning at their vowel, and in turn compete even more effectively with *mint* for the [t]. A direct boost for such words was not necessary since simply removing their competitors achieved the same result. (Note that with a correctly aligned word, or a weak following syllable, the presence of strongly activated candidates for immediately following input will help detection of the target word by competing with those candidates that overlap with the target.)

The PWC seems therefore to be a powerful and general constraint on segmentation which can exploit, in essentially one unified and very simple manner, boundary cues provided both by silence and by metrical structure. As this simulation has shown, the PWC disfavors not only words which are misaligned with a segmentation cue external to them (e.g., *apple* in [fæpəl], because of the cue of silence before the [f]), but also words which appear impossible given a segmentation cue within them (e.g. *mint* in [mɪntaup]).

#### b. *Vroomen and de Gelder (1995)*

The boundary cues which are relevant to segmentation are highly language-specific. Therefore the operation of the Metrical Segmentation Strategy differs across languages. In a stress rhythm language like English, the onsets of strong syllables constitute clear boundaries, but the onsets of weak syllables do not. In a language like French, with a syllabic rhythm, all syllable boundaries should be clear. The PWC should be implemented in such a way that it would operate upon the set of specific cues appropriate for a given language.

The metrical stress pattern effect observed in English has, as it happens, also been found in another stress-timed language: Dutch. In addition, effects of competition between lexical hypotheses similar to those found in English have been observed (Vroomen & de Gelder, 1995; Vroomen et al., 1996). We next ask whether the ability of the PWC in Shortlist to simulate the detailed pattern of the experimental results from English can generalise to a different language, using a different lexicon and set of experimental materials from a different laboratory paradigm. Vroomen and de Gelder (1995) studied



**FIG. 5.** Mean target activation levels for the materials from Vroomen and de Gelder (1995). The model was run on a 20004-word Dutch lexicon, using the Possible-Word Constraint. The mean activation functions are shown for target words in StrongWeak strings with no second syllable competitors (e.g., *melk* in [mɛlkəm]), in StrongStrong strings with few second syllable competitors (e.g., *melk* in [mɛlkøm]), and in StrongStrong strings with many second syllable competitors (e.g., *melk* in [mɛlka:m]). The activation functions are aligned with the Consonant/Vowel (C/V) structure of the target-bearing strings and three following silent segments (“[’”), such that, for example, the [k] of *melk* is aligned with the second C of the medial consonant cluster.

the relation between stress pattern and number of competitors in Dutch using a cross-modal priming task. Because of the constraints of Dutch, their experiment had only three conditions: a SW control with no second syllable competitors [mɛlkəm]; SS with few second syllable competitors [mɛlkøm]; and SS with many competitors [mɛlka:m]. Their results with this task, however, were directly comparable to the English word-spotting results; they found that the availability of more competitors increased the size of the metrical effect.

Figure 5 shows Shortlist simulations of their experiment using a Dutch lexicon of 20004 words (derived from the CELEX Dutch lexicon; Burnage, 1990) and the same parameters as for the English simulations. Over the last four segment positions of the simulation all three conditions differ from each other significantly, showing that the size of the metrical effect increases as the number of competitors increases, exactly as was observed in their study.

The PWC in fact has two roles to play in the Dutch simulations. The first PWC effect has already been discussed: a target in a SS string is penalized immediately after its offset (on the vowel of the second syllable), when it becomes clear that there is a syllable boundary before its final phoneme (e.g., before the [k] in [mɛlka:m]) with which it is misaligned. But the PWC also influences the activation of targets in SW strings. Notice that the mean activation of these targets falls slightly in the second syllable, and rises considerably after the onset of silence. This pattern is due to embedded SW words (like *melken*, to milk). The Dutch suffix *-en*, used to mark infinitives, and plurality

in both nouns and verbs, is standardly pronounced as a schwa, that is, with the final [n] deleted. Inflected adjectives have the suffix *-e* which is also pronounced schwa. These SW words are the Dutch equivalent of the English SW words, like *jumper* in [ɔ̃ʎ ʌmpəv], which, as we described earlier, can be detected by computational models but are not spotted by human listeners. Without the PWC, such words would attain a higher level of activation than the intended target words, *jump* or *melk*. The PWC, however, penalizes these longer embedded words, because they are misaligned with the clear syllable boundaries at the offset of the SW strings (there is only a consonant, such as the final [m] in [mɛlkəm], between the edge of *melken* and the clear boundary cued by silence). Before the end of the second syllable, words like *melken* compete with *melk*, keeping activation of the shorter word down. But after silence has been detected, the longer words are penalized, and the shorter words are free to dominate the activation pattern. In line with the human data, where the shorter rather than the longer words are detected in SW strings, the present simulations thus correctly predict a preference for the shorter words. The PWC account of segmentation and competition can therefore readily generalize to other languages.

### c. Cutler and Norris (1988)

Simulations based on the materials used in Cutler and Norris (1988; Experiment 3) using the PWC have also been carried out; no separate figure is given here since it would merely recapitulate the clear difference between SS and SW which can be seen in Figs. 4 and 5. As in the Norris et al. (1995) and the Vroomen and de Gelder (1995) simulations, the activation levels of target words in SS strings (e.g., *mint* in [mɪntɛɪf]) are considerably lower than those of target words in SW strings (e.g., *mint* in [mɪntəf]).

Even though some Cutler and Norris (1988) SW materials such as [dʒʌmpəv] contained embedded words like *jumper*, listeners did not detect the longer word; they spotted *jump* instead. The simulations accurately capture this performance on these materials, just as in Dutch. At the offset of the SW string, the longer (mismatching) words are penalized by the PWC, and, as in Fig. 5, the activation of the target words rise. (Note that no such effect is visible in the SW curves in Fig. 4, because the materials in the Norris et al. (1995) study were carefully selected to avoid embeddings such as *jumper*.)

The overall difference in activation of targets in SS and SW strings confirms that the basic metrical effect which motivates the MSS can be simulated by the PWC in Shortlist without any further specifically metrical mechanism. The result is simulated over a range of different English experimental materials, and with an equally comprehensive set of experimental materials in Dutch (and in consequence using two separate, but realistically sized, lexicons). All that is needed is that strong syllable boundaries in the input be considered an explicit boundary cue in the sense exploited by the PWC; this renders unnecessary any especial implementation of the MSS such as that reported in Norris et al. (1995). The PWC alone can effectively implement the MSS

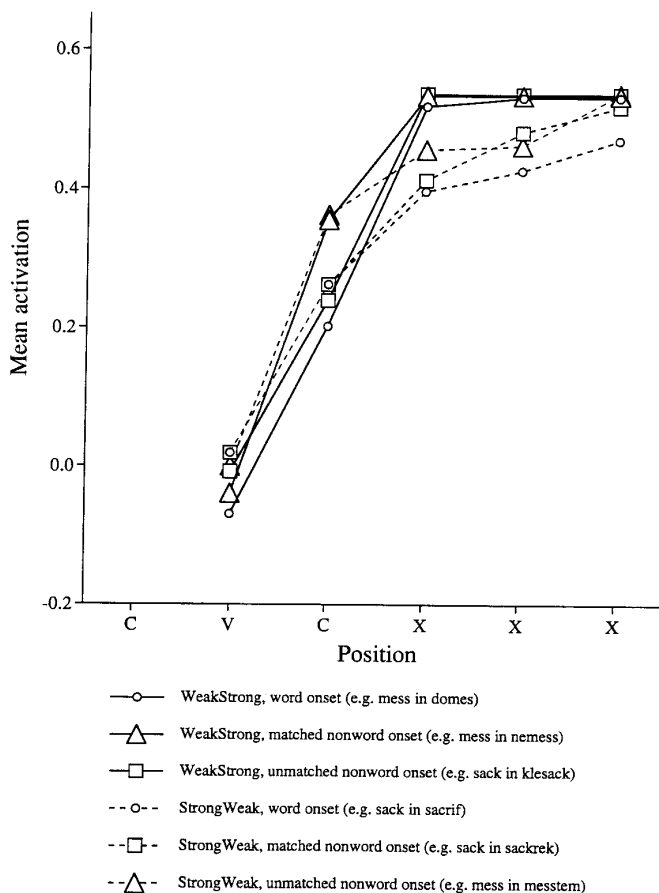
just by being able to exploit metrical structure in the input, but it has the added power of being able to do this in exactly the same way that it uses other cues to syllable boundaries (such as that provided by silence in the current experiments) to assist in segmentation.

*d. McQueen, Norris, and Cutler (1994)*

An important component of Norris et al.'s (1995) argument that the MSS involved both a boost and a penalty component came from the word spotting experiment by McQueen et al. (1994). They found that subjects were better at detecting a word when it formed the second syllable of a weak–strong string (WS, e.g., *mess* in [nəmɛs]) than when it was the first syllable of a strong–weak string (SW, e.g., [mɛstəm]). This difference was explained by assuming that the boost component of the MSS should only trigger when the strong syllable was in non-initial position. In initial position there is no reason to favor a strong onset over a weak onset as silence is normally an unambiguous cue to a word onset. The advantage of targets in WS strings then emerges because such targets are boosted at their onset and have a clear boundary at their offset. Targets in SW strings, however, have only a clear boundary at their onset and have no segmentation cues at their offset.

As mentioned earlier, the McQueen et al. results are also important because they provide one of the clearest demonstrations of competition in spoken word recognition: detection of targets at the end of nonsense strings was harder when the string was the beginning of a longer real word than when the string could not be continued to form a real word (e.g., *mess* in [dɛmɛs] vs [nəmɛs]). Similarly, words at the beginning of strings were harder if the strings could be continued to form words than if they could not (e.g., *sack* in [sækrɛf] vs [sækræk]). These competition effects are a natural property of a competition-based model like Shortlist and emerge in all versions of the model. It is important to ascertain that these competition effects, which are robustly attested in experimental situations, are not compromised by the addition of the PWC mechanism to Shortlist.

Figure 6 shows simulations using the stimuli from McQueen et al. (1994) produced by the PWC in Shortlist, with no additional MSS components. The model successfully simulates both the competition effects and the stress effects observed by McQueen et al. Figure 6 shows that activation levels are higher for target words in WS strings than targets in SW strings (compare the triangles on solid lines, e.g., *mess* in [nəmɛs], with the triangles on dashed lines, e.g., *mess* in [mɛstəm], at the two positions after offset of the target word). The competition effects in these data are also preserved with the PWC implementation: target activation levels are higher in strings which cannot be continued to form words than in strings which are the onsets of longer words. As in the human data, this lexical competition effect is larger in WS strings (circles versus triangles on solid lines, particularly at the last segment of the target word) than in SW strings (circles versus squares on dashed lines at the final two positions), for the reasons discussed in the introduction. In WS



**FIG. 6.** Mean activation levels of target words and embedding words for the items from McQueen et al. (1994) in Shortlist. The model was run on a 26455-word English lexicon, using the Possible-Word Constraint. The solid lines show activation of targets embedded in WeakStrong strings: in word onsets (circles; e.g., *mess* in [dɒməs], the onset of *domestic*), in nonword onsets matched to word onsets (triangles; e.g., *mess* in [nəməs]), and in unmatched nonword onsets (actually matched to targets in StrongWeak strings: squares; e.g., *sack* in [kləsæk]). Dashed lines show activation of targets embedded in StrongWeak strings: in word onsets (circles; e.g., *sack* in [sækrəf], the onset of *sacrifice*), in nonword onsets matched to word onsets (squares; e.g., *sack* in [sækræk]), and in unmatched nonword onsets (actually matched to targets in WeakStrong strings: triangles; e.g., *mess* in [məstəm]). The activation functions are aligned relative to the Consonant/Vowel (C/V) structure of the target words (CVC *mess* or *sack*), followed by three segments (X), which correspond either to silent segments (after WeakStrong strings) or to the second syllables of the StrongWeak strings.

strings, the difference in activation levels for targets in word onsets and nonword onsets was significant on the penultimate segment of the target word ( $t(17) = 5.67, p < .001$  two-tailed, as are the other tests below), the final segment of the target ( $t(17) = 4.04, p < .005$ ), and the first segment after

target word offset ( $t(17) = 3.15, p < .01$ ). In SW strings, this difference only begins to appear on the second segment after target word offset, where it was marginally significant ( $t(17) = 1.86, p = .08$ ), and the difference only reaches significance on the third segment after target offset ( $t(17) = 2.64, p < .05$ ).

The PWC here produces the same metrical effects as the Norris et al. (1995) instantiation of the MSS, but for different reasons. In a WS string there are three clear boundaries: silence before the string, onset of the strong syllable, and silence at the end. In a SW string only silence provides clear boundary cues, at either end. In WS strings the PWC will penalize competitors that are misaligned with silence or with the strong syllable (e.g., *amend* in the string [nəməs]). In SW strings only competitors misaligned with the silence will be penalized. Therefore, with the PWC, the WS advantage comes from the elimination of competitors at the internal strong onset rather than, as in Norris et al.'s (1995) implementation, from directly boosting candidates that begin at that strong syllable onset. However, this difference between SW and WS words observed in the present simulation is only marginally significant. Comparing the WS Nonword strings such as [nəməs] with the SW Nonword strings such as [mɛstəm] the results are: two segments after the last phoneme of the target word:  $t(17) = 2.05, p = .056$ , two-tailed; three segments after the last phoneme of the target word:  $t(17) = 2.00, p = .062$ , two-tailed. Although the effect is larger (and significant) in the comparison between the WS Nonword strings such as [kləsæk] and the SW Nonword strings such as [sækræk], this difference is in part due to lexical competition from longer words in the SW nonwords (e.g., from *sacrifice*).

These simulations have therefore produced a powerful and surprising result: not only can the PWC capture the effects which motivated its design, it can also, given only specification of strong-syllable locations in the input, completely capture the robust metrical effects established now in so many experiments in English and Dutch. The metrical effects are simulated by the—conceptually simple—PWC without any further specific metrical mechanism being necessary. Moreover, the operation of the PWC accurately captures the precise manner in which the metrical effects interact with competition effects, without in any way doing injustice to the robust competition effects themselves.

What counts as a known boundary for purposes of the PWC's operation can be silence or it can be a strong syllable; but further options also present themselves. For instance, as we described in the introduction, phonotactic sequence constraints can also rule out or force syllable boundaries. In the following section we explore the operation of the PWC given information about phonotactically signalled boundary location.

#### 4. Phonotactic Constraints

McQueen (submitted) showed that Dutch listeners found it harder to detect words in bisyllabic nonsense strings when the words were misaligned than when they were aligned with a syllable boundary. These syllable boundaries



were determined by phonotactic constraints. For example, *rok* (skirt) was more difficult in [fi.drøk] (the [d] must be syllable-initial) than in [fim.røk] (a [mr] cluster is illegal). Large alignment effects were found in four different conditions: in SS strings where the target was preceded by contextual information (e.g., [fim.røk] and [fi.drøk]); in similar WS strings (e.g., [fəm.røk] and [fə.drøk]); in SS strings where the target was followed by contextual information (e.g., *vel*, *skin*, in [vəl.brul] and [vəlm.rul]); and in similar SW strings (e.g., [vəl.brəl] and [vəlm.rəl]).

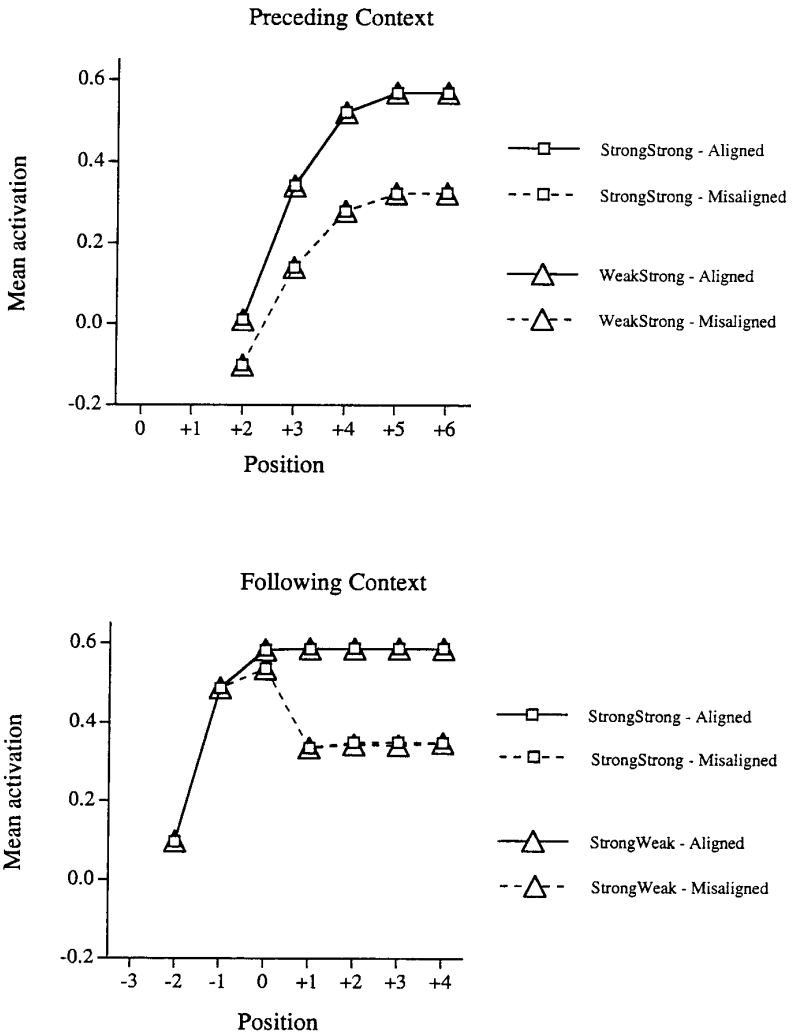
Figure 7 shows that the PWC in Shortlist can simulate these results. The model was run with the same Dutch lexicon and the same parameters as those used in the simulations of the Vroomen and de Gelder (1995) results, described above. Phonotactically determined boundaries were marked in the input. The activation levels of targets which were aligned with phonotactic boundaries are considerably higher than those of targets which were misaligned with boundaries, both for preceding and following context. Mean activation levels are almost identical across stress patterns. This is because items were very closely matched, with only the vowels in the context varying over stress patterns. In all cases, misaligned targets are penalized by the PWC, because there is a non-vocalic segment (an impossible word) between the edge of the target word and a clear syllable boundary.

The PWC seems therefore to be able to exploit boundary information cued not only from silence and metrical structure, but also from phonotactics. Note that as with metrical cues, phonotactic legality is language-specific. The item [fidrøk], for example, has a boundary cue before the [d] in Dutch (since stops are devoiced in syllable-final position, a voiced stop must be syllable-initial), but this cue would not exist in English. The PWC should be implemented in such a way that it would operate upon the set of specific cues appropriate for a given language.

Note finally that evidence for candidates violating the PWC has been reduced in our implementation by a constant proportion (by half). That is, bottom-up evidence is reduced by the same constant proportion regardless of the nature of the syllable boundary that leads to a violation of the PWC. A more realistic simulation, however, might be achieved by modulating the degree of reduction according to the strength of the evidence in favor of a boundary. For example, silence might be considered to be a particularly strong cue which should have a larger effect than, say, a phonotactically-determined boundary, which in turn might have a larger effect than a strong syllable onset. These considerations should be addressed in further investigations.

## 5. Continuous Input

The preceding sections have demonstrated that the PWC's operation manifests considerable and somewhat surprising generality, capturing with precision a number of different effects in simulations with the full experimental materials from a wide range of experiments in two languages. This powerful achievement motivated us finally to explore its performance with a more



**FIG. 7.** Mean target activation levels for the materials from McQueen (submitted). The model was run on a 20004-word Dutch lexicon, using the Possible-Word Constraint. The upper panel shows activation functions for target words in the second syllable of both StrongStrong and WeakStrong strings, preceded by contextual information, such that the target was either aligned with a phonotactic boundary (e.g., *rok*, *skirt*, in [fɪm.rɔk] and [fəɪm.rɔk]) or misaligned with a phonotactic boundary (e.g., *rok* in [fi.drɔk] and [fə.drɔk]). The lower panel shows activation functions for target words in the first syllable of both StrongStrong and StrongWeak strings, followed by contextual information, such that the target was again either aligned with a phonotactic boundary (e.g., *vel*, *skin*, in [vɛl.brul] and [vɛl.brəl]) or misaligned with a phonotactic boundary (e.g., *vel* in [vɛlm.rul] and [vɛlm.rəl]). The activation functions are aligned relative to the last/first phoneme of the context (0). Thus, for targets with preceding context, 0 is, for example, either the [m] or the [d] in [fɪm.rɔk] and [fi.drɔk], while for targets with following context, 0 is, for example, either the [b] or the [m] in [vɛl.brul] and [vɛlm.rul].

substantial input, more closely approximating real speech than an experimental situation. We examined the performance of the PWC in Shortlist on a large sample of continuous input. This input was selected more or less at random from one of the authors' bookshelves: the first eight paragraphs (1034 words) of Pinker (1994).

### *Method*

The text was first transcribed phonetically. This was done by looking up the phonetic transcriptions of each orthographic word in a large machine-readable dictionary. Those words not found in the dictionary were transcribed by hand. Automatic transcription has one major limitation: for words with multiple pronunciations there is no simple means of choosing between these alternatives. In particular, given that the PWC operates on metrical information, it was important to make realistic choices on the pronunciation of the function words in the text, most of which have pronunciations with both strong and weak vowels (e.g., *and* as either [ænd] or [ənd]). A native speaker of British English, naive as to the purpose of the recording, was therefore asked to read the eight paragraphs. The second and fourth authors then independently transcribed all the function words in the speaker's recording. There was 75% agreement on which words had been produced with reduced vowels. The 245 words that were agreed on were assigned their reduced pronunciations in the transcription. All other function words were assigned their full-vowel transcriptions.

In addition, two segments of silence were placed at the end of each clause in the transcription, as defined by the punctuation of the orthographic text, such that they were placed at every period, comma, semi-colon, colon, question mark and hyphen. The two transcribers agreed that the speaker had indeed produced intonational boundaries at all punctuation marks. In addition, they detected some further pauses or other intonationally marked boundaries in the recording which were not aligned with punctuation marks in the text. Any such boundary that was noted by both transcribers was also marked in the transcription with two segments of silence. The Shortlist lexicon was expanded to include the 257 words which appeared in the text but were not in the lexicon, giving a total lexicon size of 26712 words.

The benefits of the PWC were examined in a range of simulations. In order to facilitate unambiguous scoring of the Shortlist output of simulations of the complete text, word activation values were examined at the end of the two segments of silence following each clause. A word that was actually in the input was scored as a correct recognition (a hit) if its activation was greater than a threshold value (0.2) at this point. A word not in the input was counted as wrongly recognized (a false alarm) if its activation was also above this threshold. Each simulation was run both with clear input, in which all input phonemes were specified correctly, and with degraded input, in which 20% of the input phonemes, selected at random, were replaced with an ambiguous phoneme. Vowels were replaced with an ambiguous vowel, [V]. In the computation of bottom-up support for lexical candidates, this ambiguous vowel mismatched with all consonants (i.e., candidates with a consonant aligned with [V] in the input had their bottom-up support reduced) but did not mismatch with any vowels (i.e., candidates with a vowel aligned with [V] had their bottom-up support neither increased nor decreased). Likewise, consonants were replaced with an ambiguous consonant, [C], which mismatched all vowels but did not mismatch with any consonants.

### *Results*

*Baseline.* The results of the simulations are shown in Table 3. The baseline condition is how well Shortlist performs with lexical competition as the only segmentation process. With clear input, 90% of the words were recognized by competition alone. There were 29 false alarms. With degraded input, performance was of course poorer. Nevertheless, even when 20% of the

TABLE 3

The Benefits of the PWC in Shortlist, Measured on Performance on the Phonetic Transcription of 1034 Words of Text Taken from Pinker (1994)

Test	Hits	Misses	Errors	False alarms
Clear input				
Competition alone	931 (90.0%)	63 (6.1%)	40 (3.9%)	29
PWC with				
Silence & phonotactic cues	943 (91.2%)	57 (5.5%)	34 (3.3%)	28
Silence, & phonotactic & metrical cues	993 (96.0%)	26 (2.5%)	15 (1.5%)	10
All word boundaries marked	998 (96.5%)	21 (2.0%)	15 (1.5%)	5
Possible resyllabifications unmarked	976 (94.4%)	41 (4.0%)	17 (1.6%)	18
Degraded input				
Competition alone	592 (57.3%)	341 (33.0%)	101 (9.8%)	137
PWC with				
Silence & phonotactic cues	624 (60.3%)	335 (32.4%)	75 (7.3%)	107
Silence, & phonotactic & metrical cues	671 (64.9%)	303 (29.3%)	60 (5.8%)	50
All word boundaries marked	685 (66.2%)	298 (28.8%)	51 (4.9%)	24
Possible resyllabifications unmarked	658 (63.6%)	318 (30.8%)	58 (5.6%)	66

*Note.* The model was tested with lexical competition alone (Competition alone) and with competition plus the Possible-Word Constraint (PWC), operating on different boundary information (Silence & phonotactic cues; Silence, & phonotactic & metrical cues; All word boundaries marked; and Possible resyllabifications unmarked). Performance is given both when the input was perfectly transcribed (clear) and when the input was degraded. See text for further details. Hit = Activation of input word > 0.2; Miss = No word with activation > 0.2 with onset aligned with that of input word; Error = Erroneous word with activation > 0.2 with onset aligned with that of input word; False alarm = Word with activation > 0.2 with onset not aligned with any input word. The hits, misses, and errors sum to the total number of input words, and are given as percentages. The number of possible false alarms is large (the total number of segments minus the total number of words in the input).

segments in the input were ambiguous, the competition process allowed more than half the words in the text (about 57%) to be recognized. There were then 137 false alarms.

*Silence and phonotactics.* How does the PWC fare relative to these baselines? We have argued that the PWC operates on likely word boundaries, cued by silence, metrical structure, and by phonotactics. In these simulations we compare the availability of different combinations of cues. Boundaries derived from phonotactic constraints can be determined algorithmically from

a phonemic transcription by applying knowledge of the possible syllable onsets and offsets in the language. In contrast to other cues to syllabification, phonotactic cues are thus independent of the precise phonetic realization of the input. In this simulation, therefore, we provided the PWC only with the clear boundaries provided by silence and by phonotactic constraints. All syllable boundaries determined by phonotactic constraints, as computed by an automatic parsing routine, were marked in the input; if a phoneme sequence yielded an illegal syllable onset or offset (as defined for English in Gimson, 1980), a boundary was marked in the appropriate location (e.g., a boundary would be placed between [m] and [r] given the string [mr] since this sequence is both an illegal onset and an illegal offset). The same procedure was applied to the degraded input for all sequences where there were no ambiguous phonemes. Where there was an ambiguous phoneme, boundaries were marked only when the phonotactic constraint was powerful enough to apply to all consonants or all vowels (e.g., a boundary would be marked between [V] and [h] in the sequence [Vh] since [h] cannot occur in any syllable coda).

Relative to the competition-only baseline, the PWC operating on silence and phonotactics showed a small benefit of 12 words in hits (almost 12% of the 103 words not recognized by competition alone), and a very small benefit in false alarms, with a decrease of only one word. Clearly, with high-quality input, silence and phonotactics provide the PWC with cues which result in only limited improvements in performance. However, larger benefits of the PWC were seen when the input was degraded. Relative to the baseline, there were 32 more correct recognitions (about 7% of the 442 words not recognized by competition alone), and 30 fewer false alarms when the PWC operated on silence and phonotactic cues.

Note that since the simulations were run on a clause-by-clause basis, the competition process alone was already provided with some segmentation information from silence. Candidates straddling intonational boundaries (beginning before the silence markers at the end of one clause and ending in the onset of the following clause) could not attain high levels of activation. But with no PWC operating, there is no means by which candidates misaligned with silence could be penalized. There are clearly few such words activated by the clear input.

These simulations therefore suggest that boundaries cued only by silence and phonotactics do not provide the PWC with sufficient information to segment accurately all the words in continuous speech. Harrington, Watson and Cooper (1989), in another computational analysis of continuous speech input, have also argued that phonotactics provide a rather weak source of information for segmentation. Only 37% of the word boundaries in their corpus could be detected on the basis of sequence constraints, and even this was only possible when the input was clearly specified. The present simulations nevertheless suggest that when boundaries cued by silence or by phonotactics are available, the PWC will act to penalize spurious lexical hypotheses, and thus improve recognition performance.

*Metrical cues.* In the next simulation, the segmentation procedure was provided with more information about likely word boundaries. We have argued that metrical information provides the PWC with further cues to likely word boundaries. Thus to the silence-cued and phonotactic boundaries we next added metrical cues, by marking all strong syllable onsets in the input. In this simulation, larger PWC benefits accrued. With clear input, 62 more words were correctly recognized than in the baseline competition-alone case (i.e., 60% of the baseline misses and errors were corrected), and there were 19 fewer false recognitions (i.e., the number of false alarms was more than halved). With degraded input, 79 more words were correctly recognized (i.e., 18% of the baseline failures were corrected), and there were 87 fewer false recognitions (the false alarm rate was thus again more than halved).

It is quite clear that metrical cues to likely word boundaries provide the PWC with a richer source of information than silence and phonotactics. Silence, as defined here, is limited to intonational boundaries, and many syllable boundaries in the text are not phonotactically marked. In contrast, the majority of words begin with strong syllables, a fact about English which motivated the original MSS (Cutler & Carter, 1987). In the Pinker (1994) text, as coded here, there are 814 strong syllables and 782 weak syllables (51% and 49% respectively). But 665 (82%) of the strong syllables are in word onset position; that is, 64% of the 1034 words begin with strong syllables. Half the syllable boundaries in the text are thus marked by the metrical cue, and the majority of these are in fact also word boundaries. Note that the PWC acts to benefit recognition performance on the basis of *all* strong syllable onsets, both those strong onsets actually at word boundaries (candidates whose edges are close to but misaligned with these boundaries will be penalized) and those strong onsets internal to polysyllabic words (e.g., erroneous candidates embedded in such words but misaligned with an internal metrically cued boundary will be penalized).

*All word boundaries marked.* The PWC could of course also make use of further (allophonic and acoustic) cues to likely word boundaries. If all word onsets were marked in the input, the listener's task would of course be easier than it is; we know that weak word-initial syllables, especially, are unlikely to be marked. However, we carried out a further simulation in which we provided the PWC with marking for all word boundaries, in order to estimate the upper limits of the benefits it can provide (especially with degraded input). The values given in Table 3 for "All word boundaries marked" are thus those where in addition to all boundaries marked by silence, and those marked by phonotactic and metrical cues (both at word edges and word internally), the onsets of all remaining (weak-initial) words were marked. With this information, and clear segmental input, 96.5% of words were correctly recognized (only 36 words out of 1034 were not recognized), and there were only 5 false alarms. With degraded input, the PWC under this condition improves the hit rate by 93 words relative to the baseline, and cuts the number of false alarms by 113 words.

The assumption that all word onsets are marked in continuous speech is clearly untenable; the simulations with all word onsets marked thus represent an upper limit on performance. In fact, performance was still less than perfect even when all boundaries were marked. This is attributable to two factors: First, our choice of threshold was somewhat arbitrary and may have been marginally too high—the single-phoneme word ‘or’ ([ɔ] in British English) just failed to reach the threshold (it consistently had an activation of 0.18). Second, when word boundary information is used to exclude impossible segmentations rather than to force segmentation at those boundaries, some portions of the input will remain completely ambiguous. For example, ‘we can’ and ‘you are,’ each with the second word reduced, were parsed as ‘weaken’ and ‘ewer,’ respectively. Such ambiguities can only be resolved either by using boundary information in a positive sense to force segmentation (in practice boundary cues before a weak syllable in a reduced version of ‘are’ or ‘can’ are unlikely) or by consulting higher level context. (Note that Norris (1994) pointed out that Shortlist could be integrated with the checking model (Norris, 1986) account of context effects on word recognition.)

As can be seen in Table 3, competition alone, with no boundaries marked and no PWC, already does a good job in solving the segmentation problem. Given the pervasiveness of boundary ambiguity, particularly with less than perfect input, there will clearly nevertheless be a need to use contextual information to arrive at a proper interpretation of the input. The assumption that no word onsets are marked in continuous speech is however also untenable: there are boundaries cued by silence, by phonotactics, by metrical structure, and by allophonic and other acoustic cues. The operation of the PWC will ensure that the input to higher level contextual processes has taken full advantage of the segmentation cues available. These simulations with continuous input show that the PWC has a clear benefit: the more boundaries that are signaled, the better performance becomes.

*Resyllabification.* So, how many word onsets are marked in continuous speech? Unfortunately, we cannot currently answer this question. Acoustic-phonetic studies (Lehiste, 1972; Nakatani & Dukes, 1977) suggest simply that some but not all onsets are marked. We have argued that silence and phonotactics may provide unambiguous sources of information about word boundary locations. It is less certain that onsets before all strong syllables will be clear. The claim that formed the basis of the MSS (Cutler & Carter, 1987; Cutler & Norris, 1988) was that syllables with full vowels will tend to have unambiguous onsets; this assumption has been made in the simulations on metrical segmentation reported here. But it is important to stress that this is a probabilistic claim, which is difficult to justify for every word in the Pinker (1994) text. Some strong onsets may well not be marked, particularly those for vowel-initial words, preceded by consonants which may resyllabify (e.g., in the Pinker (1994) text, the [t] may resyllabify in *blunt instruments* to produce *blun tinstruments*). As Kahn (1980) has argued, however, resyllabification in English is not complete; resyllabified stops are not aspirated. It is thus perhaps reasonable to

assume that when resyllabification takes place in contexts such as these, no clear word boundary will be marked (for example, there would neither be a clear boundary before the [ɪ] of *instruments* nor before the resyllabified [t]).

A final set of simulations were therefore run in which the onsets of the 73 words in the text beginning with full vowels and occurring at possible resyllabification sites (such as *instruments*) were not marked. These simulations were identical to those where cues from silence, phonotactics and metrical information were provided, except that these 73 strong word-onsets were unmarked. As one would predict, the PWC acts to produce a level of performance between that which occurs when only boundaries cued by silence and phonotactically determined boundaries are marked and that which occurs when all strong onsets are marked. But note that the decrement in performance on correct recognitions is only 17 words with clear input and only 13 words with degraded input, that is, in both cases, far fewer than 73 words. It is clearly not the case that the model simply fails to recognize all the words at unmarked boundary locations. Competition alone is often sufficient for a correct parse to be obtained.

It is important to stress that these simulations are approximations, and tend to exaggerate the problems of resyllabification for the PWC. Resyllabification is an optional process. Some words beginning with full vowels may well be unambiguously marked at their onsets by an acoustic cue. One such cue, which can occur before both strong and weak word-initial vowels, is glottalization (Dilley, Shattuck-Hufnagel & Ostendorf, 1996). In a study of the speech of five radio-news broadcasters, Dilley et al. (1996) report that glottalization before vowel-initial words was most common when the words occurred in phrase-initial position. But evidence of glottalization was also obtained for non-phrase-initial words, with the likelihood of glottalization tending to increase from words with weak vowels to those with strong vowels in unaccented syllables to those with strong vowels in accented syllables. There was also considerable between-speaker variability (ranging from an overall rate of only 13% in one speaker to 44% in another).

These results show that even for this cue alone it is impossible to predict how often words beginning with either strong or weak vowels will have their onsets marked. As we have argued above, there are no fully reliable cues to word boundaries. But Dilley et al.'s (1996) results also suggest that vowel-initial words beginning with strong syllables are more likely to be marked by glottalization than are those beginning with weak syllables. Since resyllabification will be blocked by glottalization, resyllabification is thus less likely before strong syllables (where it would be damaging to the operation of the PWC) than before weak syllables (where the PWC is not metrically cued). Note also that when glottalization does occur before a weak syllable, the PWC would be able to use that cue to improve recognition performance.

### *Discussion*

These simulations show that the PWC improves recognition performance in continuous speech. The PWC tends to improve performance no matter how



a likely word boundary is marked, but metrical cues appear to provide the richest source of such information. Although it is uncertain how many likely word boundaries are actually marked in normal speech, it is clear that as this number increases, so too will the benefits of the PWC increase.

Given the nature of its operation, the main benefit of the PWC is in filtering out false-alarms that fail to line up with possible boundaries in the input. Its effect on hit-rate is an indirect consequence of removing these unwanted competitors. In general, the benefits of the PWC in terms of absolute improvement in performance were stronger given degraded input than with fully specified input. This is simply because there was greater scope for improvement (the baseline was 90% correct recognitions with clear input). Since it is reasonable to suppose that the input for lexical access is not a perfect phonemic transcription, the degraded input simulations may provide a more realistic estimate of the benefits of the PWC than do the clear input simulations. Nevertheless, in both cases there is a clear tendency towards more hits and fewer false alarms as the PWC is provided with more boundary cues. These simulations suggest that the efficiency and success of the listener's task of continuous speech recognition can be enhanced by two closely linked factors: sensitivity to cues to likely word boundaries in the speech signal and a knowledge of the acceptability of sections of speech material as possible words in the language.

#### GENERAL DISCUSSION

This paper has argued for a powerful and general constraint on the segmentation of continuous speech input into its component words. We have presented two experiments which show clearly that listeners find it difficult to recognize words embedded in a context which could not by itself constitute a word; detection of *apple*, for example, proved much more difficult in *fapple* than in *vuffapple*. We have argued, on the basis of this finding, that prelexical segmentation of speech by human listeners is subject to a Possible-Word Constraint (PWC). We propose that knowledge about what constitutes a possible word in the listener's native vocabulary is used on-line, during spoken word recognition, to modulate the activation of lexical hypotheses. A word boundary after the [f] in [fæpəl] is highly improbable, and *apple* thus difficult to detect, because *f* is not a possible word of English. In contrast, *apple* in [vʌfæpəl] is easier to detect because the syllable [vʌf] could conceivably be a word, and therefore a lexical boundary after the [f] is possible.

The operation of the PWC, we have claimed, depends upon segmentation cues in the input; these cues provide, in varying ways, information about syllable boundaries in the input, and the PWC disfavors candidate words which are misaligned with these boundaries. A word is considered to be misaligned when there are no vocalic segments (and hence no possible syllables, and in turn no possible words) between the edge of that word and such a boundary. In the current experiments, the effective cue was simply provided by silence: there is clearly a syllable boundary before the [f] in [fæpəl]. Thus

listeners can efficiently exploit the segmentation cue provided by silence during the recognition of continuous speech. But many other kinds of segmentation cues can be exploited in the same way; the PWC by no means depends on the availability of silent intervals. It allows, for instance, the exploitation of phonotactic sequencing constraints in just the same way; and it also allows the exploitation of metrical cues to boundary location.

### *The PWC and Competition*

The PWC does not use boundary cues deterministically to force segmentation prelexically. Instead, it constrains the activation and competition process by which spoken-word recognition occurs, to disfavor candidate words not aligned with these boundaries. Competition, as we have argued, provides a mechanism both for the recognition of individual words, and for the segmentation of continuous speech. In the introduction we reviewed the large body of research which has provided evidence of activation of multiple lexical hypotheses, and competition among them; it is clear that human spoken-word recognition involves such a process of competition.

Indeed, all current models of continuous speech recognition are based on a competition mechanism (see McQueen et al., 1995). Older models of continuous speech recognition, which postulated that words were recognized in strictly sequential order (Cole & Jakimik, 1978, 1980; Marslen-Wilson & Welsh, 1978), have now been abandoned. Strictly sequential models are unable to use later-arriving information (such as the [θ] in *met a fourth time*) to influence perceptual decisions about earlier words (such as *metaphor*). More recent models, in which competition is considered only to operate at a perceptual decision stage (e.g., Marslen-Wilson, 1993) also lack an effective means by which information beginning at different points in time can be used to settle on an optimal interpretation of the input. In contrast, competition between candidate words beginning at different points in the input, as in Shortlist, provides exactly this kind of mechanism (see McQueen et al., 1995, for further discussion). It was therefore both natural and straightforward to implement the PWC as a bias which influences the optimal interpretation of the input that the competition process settles on.

The PWC has been implemented as a bias in the Shortlist's competition process. Candidate words which are misaligned with clear syllable boundaries are penalized: their activation values are halved. Simulations have shown that this computational implementation of the PWC accounts well for the current experimental data, in that activation values are lower for words embedded in contexts (either preceding or following the word) which could not themselves constitute words.

Further simulations have shown that the PWC also provides an account of data from previous studies on speech segmentation (Cutler & Norris, 1988; McQueen et al., 1994; Norris et al., 1995; Vroomen & de Gelder, 1995) which have demonstrated that listeners exploit rhythmic structure in segmenting continuous speech. The Metrical Segmentation Strategy (MSS)

formulates the claim that in stress-timed languages like English and Dutch, listeners postulate word boundaries at the onsets of strong syllables. Strong syllables tend to provide cues to clear syllable boundaries. The PWC can therefore act to penalize words which are misaligned with these metrically determined boundaries (in just the same way as it acts on the cues provided by silence).

An important point to note about this account is that the PWC can only explain these data if the prelexical analysis of the speech signal provides the metrical cues specified by the MSS (in this case the location of strong syllable onsets). The marking of strong syllables is assumed to be an important part of the initial analysis of the input in languages like English and Dutch, and was provided in the current simulations as part of the input to Shortlist (just as it was provided in the simulations reported by Norris et al. [1995]). As we pointed out in discussing our simulations with the Pinker (1994) text, there may well be some strong syllable onsets which are not clear in continuous speech, due to processes such as resyllabification. But the PWC does not require that all onsets be marked. The claim is that when an onset is marked (and we would argue that most strong syllables will have unambiguous onsets) the PWC can use this information. When an onset is unmarked, competition is still available to provide an optimal interpretation of the input.

Given metrically-cued boundary locations, the PWC provides an economical implementation of the MSS, and renders redundant the alternative (and in fact more unwieldy) implementation of the MSS reported by Norris et al. (1995). The simple PWC mechanism of penalizing words which are aligned with positions that are unlikely to be word-boundaries (when there is non-syllabic speech material between such a position and a clear syllable boundary) thus not only accounts for the present data but also instantiates the MSS as motivated by numerous previous studies.

If provided with information about points of phonetic sequencing illegality, the PWC can also account for experimental results showing listener sensitivity to whether words are aligned or misaligned with syllable boundaries specified by phonotactic constraints (McQueen, submitted). Finally, the simulations based on the text taken from Pinker (1994), and those with the phrase *met a fourf time*, showed how the PWC allows more optimal interpretations of continuous speech to be obtained in a competition model. The PWC thus provides a simple but powerful account of a wide range of factors affecting the segmentation of continuous speech.

Note that the simplicity of the PWC's operation in no way denies the variety and complexity of the information it exploits. Of course it is no trivial matter to incorporate phonotactic constraint information, or language-specific metrical cues, into the initial analysis of the signal which is used as input to the word recognition process. There is abundant evidence that metrical information and phonotactic information is exploited by listeners in parsing speech, and any complete model of word recognition should be able to account

for such evidence. The important feature of the present approach is that all these sources of information are exploited indirectly, by the bias in the competition process via which the PWC has been implemented. This allows such information to be used in the word recognition process, while preserving the essential structure of this process as one of competition between candidate words.

### *Possible Words*

The PWC's operation is crucially sensitive to the cues—silence, metrical structure, phonetic sequence—present in the input. But further, the PWC distinguishes between types of phonemes: some are possible words, some are not. What currently counts as a possible word is any stretch of speech containing at least a vowel. Thus, effectively, consonants are subject to different processing constraints than vowels. Such a difference is indeed supported by recent evidence showing processing differences between vowels and consonants in human spoken-word processing (van Ooijen, 1994, 1996).

This vowel-based definition of the PWC will undoubtedly require further elaboration. If our general characterization of the PWC is correct, then it should be the case that the size and type of unit that will constitute a possible word will vary across languages. For example, as we mentioned in the introduction, some languages (such as Lardil) have no words with fewer than two syllables. If what counts for the segmentation procedure is, as we have argued, the viability of a stretch of speech as a possible word of a language, then our current formulation of the PWC should prove inadequate for languages like Lardil. A cross-linguistic analysis of segmentation performance is required to test the generality of the PWC claim.

The present formulation nevertheless appears to provide a good account of the segmentation of languages like English and Dutch. This is because what matters crucially for the operation of the PWC is what *fails* the constraint. The available data can all be accounted for by a process in which single consonants are not possible words. It may be the case that future research will show that longer stretches of speech are also impossible words in particular languages, and that they too will fail the PWC, but single consonants would still remain as impossible words.

### *Syllabic Information*

The PWC is cued by information specifying the location of syllable boundaries. Furthermore, the constraint on what constitutes a possible word is essentially a syllabic constraint (a stretch of speech without a vowel is non-syllabic). The PWC therefore reflects the very important role of the syllable in the processing of phonological structure. Syllabic information has been proposed to play a role in speech recognition in many different ways. The strongest view, represented by Mehler and his colleagues (e.g., Mehler, 1981; Seguí, 1984), is that the syllable is primarily a unit of classification. That is,

the input is classified into syllables (rather than, for example, phonemes), and it is this syllabic representation that is used to access the lexicon. An alternative view, characterized by the MSS, is that syllables, or other rhythmic units, primarily play a role in segmentation. So, in English, onsets of strong syllables indicate the likely locations of word boundaries. The syllables themselves are not units of classification and the representations used for lexical access therefore do not have to be represented as strings of syllables.

The PWC represents a modification of the segmentation view. Syllable boundaries are used indirectly to help segmentation, but the actual operation of the PWC indicates locations where word boundaries *cannot* occur rather than where they should occur. Apart from silence, which will be important in all languages, the kind of boundaries that will be used in a given language will be determined by the metrical properties of the language as specified by the MSS, and by the phonotactic cues specific to that language. The PWC emerges therefore as a universal principle of segmentation that is modulated in language-specific ways. Like the MSS, the PWC depends crucially on the ability to detect syllable boundaries (possible word boundaries) but is not dependent on the ability to classify syllables. (The version of Shortlist used to implement the PWC continues to perform lexical access on the basis of phonemic representations of the input.)

Two recent studies have suggested that listeners are sensitive to syllabic structure in English input. Bruck, Treiman, and Caravolas (1995) required listeners to decide whether two nonwords shared sounds; the task could be performed more rapidly when the nonwords shared a syllable (e.g., [kɪpæst] and [kɪpæld]) than when they did not (e.g., [flɪgmɪl] and [flɪkboz]). Although this result suggests that listeners can process syllabic information, it does not require (as the authors themselves point out) the on-line ability to classify speech input into syllables. The task may well call upon phonological memory processes beyond those used in normal speech processing. Note that in other tasks drawing upon phonological memory, English-speaking subjects fail to show syllabification effects which do appear in the responses of French-speaking subjects. Thus Peretz, Lussier, and B eland (1996) found that word stem completion responses by French-speaking subjects were sensitive to the syllable structure of the prompt (given the prompt BA- subjects were more likely to produce a word with an open first syllable, given the prompt BAL- they were more likely to produce a word with a closed first syllable), but responses by English-speaking subjects showed no such effect.

Finney, Protopapas, and Eimas (1996) have further shown that listeners can apparently use syllabic information to cue them to the location of phoneme targets in a phoneme detection task. Their experiment directly replicated the study of Pallier et al. (1993), who showed such effects in French and Spanish. Listeners more rapidly detected targets occurring in a syllabic coda when most of the targets in the experiment occurred in coda position; when most of the targets in the experiment occurred in a syllabic onset, however, listeners more rapidly detected targets in onset than in coda position. Finney et al.

observed this effect in English words with strong second syllables (e.g., *segmental*, with [g] as target in coda position; *seclude* with [k] as target in onset position). This result, in fact, follows directly from the MSS as instantiated for English; since the crucial syllable boundary is at the onset of a strong syllable, listeners can learn to focus attention on a location immediately before or after the clear syllable boundary. Interestingly, as further predicted by the MSS, the effect did not replicate in words with strong first syllables and weak second syllables (e.g., *juggler* and *secret*); in such words, the syllable boundary (and hence any manipulation of coda and onset position) is predicted by the MSS to be unclear. These results thus again do not require that the speech input be classified into syllables.

### *Resyllabification*

The PWC therefore makes use of syllabic information, but it does not depend on a syllable classification procedure. It nevertheless capitalizes on the tendency for word boundaries and syllable boundaries to be highly correlated. What will therefore happen when, due to processes such as resyllabification, word boundaries and syllable boundaries are not coincident? First, it is important to stress that in English, resyllabification is not a particularly common process: 2.56 million word tokens in a corpus of 17.9 million words (in the CELEX database, Burnage, 1990) begin with full vowels and 6.40 million tokens end with consonants which could resyllabify. On the admittedly oversimple assumption that every word was equally likely to follow every other word in the corpus, these numbers provide the estimate that about 5% of word boundaries in the corpus could involve resyllabification. In the extract from Pinker (1994) used in our simulations, it was judged that 73 out of 1034 word onsets (7.1%) could involve resyllabification. So the maximum number of possible resyllabifications in continuous English speech is likely to be rather small. Add to this the fact that resyllabification is an optional process (the speaker does not have to resyllabify at all of these possible sites), and the size of the problem of resyllabification for the PWC shrinks further.

But how might the PWC deal with resyllabification when it does occur? As already noted, resyllabification is not complete. A resyllabified consonant, such as the [k] in *bake it* does not appear to be identical to a normal syllable-initial consonant, such as the [k] in *bay kit*. In other words, *bake it* and *bay kit* are not fully homophonous, even if the [k] resyllabifies, since the [k] will not be aspirated (Kahn, 1980). This acoustic difference may be sufficient to prevent a strong onset being signaled before the [k]. Listeners are certainly sensitive to the distinction between aspirated and unaspirated stops (Christie, 1974). Note that we are only concerned here with word onsets before full vowels, since it is only at such locations that the PWC may be misled by metrical information (no boundaries are signaled before weak syllables). As we saw in the simulations based on the Pinker (1994) text, if onsets before strong syllables at possible resyllabification sites are left unmarked, many

of the words at these sites can still be recognized using competition alone. Alternatively, it may be the case that the speaker provides the listener with acoustic cues that resyllabification has occurred. If this were the case, a strong onset could be correctly marked after all, before the [ɪ] in *bake it*, for example. Both the acoustic details of resyllabified segments, and how the recognition system might use such information, remain to be determined more completely.

Ironically perhaps, notorious examples of resyllabification occur in French, which is a syllable-timed language with clear syllabification, so that all syllables should be relatively clearly marked in the signal. In French the liaison process allows resyllabification across word boundaries, for example: *petit éléphant*, can be realized as [pə.ti.te.le.fã] (where a period marks a syllable boundary). Liaison consists of two processes. In the first process, a word's latent final consonant, which is not normally pronounced (*petit* is usually realized as [pə.ti]), surfaces in the context of a following word beginning with a vowel. Second, the surfaced latent consonant resyllabifies; thus *éléphant*, which of course usually does not begin with a [t], receives the resyllabified onset.

The resulting output [pə.ti.te.le.fã] offers no barrier to the recognition of *petit*, because this word can be identified in exactly the same manner as if the latent consonant had not surfaced. There is a potential problem, however, in the recognition of *éléphant*, which now starts with a syllable beginning with the latent consonant. According to the PWC, *éléphant* should be penalized because it is misaligned with the syllable boundary before the [t]. However, the liaison process is highly systematic in a way that should allow the PWC penalty to be disabled in contexts where liaison is likely. A simple solution would be to turn off the PWC when the first word was identified and the initial consonant of the following syllable was that word's latent consonant. The importance of lexical information in helping segmentation involving liaison has been demonstrated in French by Dejean de la Bâtie and Bradley (1995). Another possibility is that there may be explicit acoustic marking of liaison (Dejean de la Bâtie, 1993) which may itself be sufficient to cause the PWC to be switched off. Note that the plausibility of disabling the PWC in liaison contexts is supported by the fact that the set of environments in which liaison can occur is very restricted in French. This is a further example of the necessity for language-specificity in implementing the input features to which the PWC is sensitive.

### Conclusion

Human listeners are extremely efficient at recognizing spoken words in continuous utterances, despite the fact that such utterances usually include spurious word-forms embedded within or across the intended words. In particular, such embeddings do not cause problems for the human listener when accepting them would leave a residue of the input which could not possibly be a word: a single consonant, for example. We have argued here that this

Possible-Word Constraint is a feature of the human spoken-word recognition system; and we have provided experimental evidence which strongly supports our proposal.

We have also presented simulations showing how the PWC can increase the effectiveness of a recognition system based solely on competition. The constraint can be simply implemented in the Shortlist model, in a way which maintains the integrity of the competition process which so much recent research has shown to be central to human word recognition. The implementation allows the model to simulate the possible-word effects in the human listening data, but it further allows it to simulate via the same mechanism a range of other experimentally attested segmentation effects based on exploitation of metrical structure and phonotactic legality in the input.

We have argued that the PWC can make use of a range of different cues to likely word boundaries; in addition to the cues provided by silence, metrical structure and phonotactics, the PWC could also exploit allophonic variation and acoustic cues to word boundaries. One obvious strength of the PWC is therefore that it offers one simple yet wide-ranging and unified account of lexical segmentation. Another strength is that the PWC, operating in the context of lexical competition, does not require that every word boundary be marked. Competition will settle on an optimal interpretation of a stretch of input which happens not to be marked by any segmentation cues. But the PWC acts to improve recognition performance as the number of likely word boundaries (cued by whatever source) increases. The PWC thus constitutes a significant further advance towards accurate modeling of the economy and efficiency of spoken-word recognition by the human listener.

## APPENDIX

Experimental materials are given by target-word length, position of context, and type of context (Possible or Impossible). The target words are shown in uppercase. The items are given in orthographic transcription.

### Monosyllabic targets

Preceding context		Following context	
Possible	Impossible	Possible	Impossible
pavACHE	vACHE	BELLshig	BELLsh
bithACT	thACT	SMELLshek	SMELLsh
yethADD	thADD	SPELLshub	SPELLsh
vutchAIM	chAIM	TELLthush	TELLth
nuthAMP	thAMP	DOLLvud	DOLLv



Preceding context		Following context	
Possible	Impossible	Possible	Impossible
lutchAPE	chAPE	OILchun	OILch
nivASH	vASH	POLEshib	POLEsh
motchAXE	chAXE	DULLfep	DULLf
davEASE	vEASE	FOOLchuv	FOOLch
tivEAST	vEAST	CHAINthiv	CHAINth
levEAT	vEAT	VEINchag	VEINch
kushEBB	shEBB	FUNchog	FUNch
pemEDGE	mEDGE	GUNchib	GUNch
futchEEL	chEEL	RUNjom	RUNdge
maffEGG	fEGG	SUNchom	SUNch
tavELF	vELF	KEYfip	KEYf
votchELM	chELM	TEAfep	TEAf
juzzELSE	zELSE	KNEEthap	KNEEthe
vutchEVE	chEVE	SEAshub	SEAsh
vuffICE	fICE	BOYsav	BOYs
gevILL	vILL	TOYnig	TOYn
rishINK	shINK	PLOUGHnev	PLOUGHn
suffOUNCE	fOUNCE	SHOEmuv	SHOEm
geffOOZE	fOOZE	ZOOthig	ZOOth

## Bisyllabic targets

Preceding context		Following context	
Possible	Impossible	Possible	Impossible
zemABBEY	mABBEY	ANCHORthim	ANCHORth
mevABSENT	vABSENT	BOTHERnem	BOTHERn
chevACTION	vACTION	COVERfum	COVERf
gezzANCIENT	zANCIENT	EAGERthib	EAGERth
guzzANGEL	zANGEL	EVERthep	EVERth
lethANGLE	thANGLE	FEATHERnuck	FEATHERn
nalANGRY	lANGRY	FOSTERnish	FOSTERn
radgeANKLE	jANKLE	GATHERmef	GATHERm
pumANXIOUS	mANXIOUS	HEATHERfak	HEATHERf
vuffAPPLE	fAPPLE	HORRORthep	HORRORth
wudgeEAGLE	jEAGLE	LAGERfek	LAGERf
foshECHO	shECHO	LEATHERnep	LEATHERn
demEFFORT	mEFFORT	MEASUREmaf	MEASUREm
kivEQUAL	vEQUAL	MIRRORfos	MIRRORf

Preceding context		Following context	
Possible	Impossible	Possible	Impossible
paffEXTRA	fEXTRA	NETHERfik	NETHERf
loshIMAGE	shIMAGE	POWDERmot	POWDERm
feshOBJECT	shOBJECT	QUIVERnal	QUIVERn
nimOINTMENT	mOINTMENT	SHIVERthig	SHIVERth
ludgeONION	jONION	SUFFERthep	SUFFERth
nithOPTION	thOPTION	SUGARthim	SUGARth
fumOVEN	mOVEN	TREMORfip	TREMORf
dalOYSTER	lOYSTER	ULCERmip	ULCERm
kesUGLY	sUGLY	USHERfav	USHERf
vishUNCLE	shUNCLE	WEATHERfud	WEATHERf

## REFERENCES

- Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 152–178). Cambridge: Cambridge University Press.
- Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, **61**, 93–125.
- Bruck, B., Treiman, R., & Caravolas, M. (1995). Role of the syllable in the processing of spoken English: Evidence from a nonword comparison task. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 469–479.
- Burnage, G. (1990). *CELEX: A guide for users*. Nijmegen, The Netherlands: CELEX.
- Chien, L. F., Lee, L. S., & Chen, K. J. (1991). An augmented chart data structure with efficient word lattice parsing scheme in speech recognition applications. *Speech Communication*, **10**, 129–144.
- Christie, W. M. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, **55**, 819–821.
- Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, **25**, 53–69.
- Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and 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.
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.
- Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, **31**, 218–236.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, **2**, 133–142.
- Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, **25**, 385–400.
- Cutler, A., Mehler, J., Norris, D. G., & Seguí, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381–410.

- 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.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, **33**, 824–844.
- Dejean de la Bâtie, B. (1993). *Word boundary ambiguity in spoken French*. Ph.D. dissertation, Monash University, Australia.
- Dejean de la Bâtie, B., & Bradley, D. C. (1995). Resolving word boundaries in spoken French: Native and non-native strategies. *Applied Psycholinguistics*, **16**, 59–81.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, **24**, 423–444.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, **97**, 1893–1904.
- Finney, S. A., Protopapas, A., & Eimas, P. D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language*, **35**, 893–909.
- Gimson, A. (1980). *An introduction to the pronunciation of English, 3rd edition*. London: Edward Arnold.
- Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and 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, and Cognition*, **18**, 1211–1238.
- Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 344–359.
- Harrington, J., Watson, G., & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, **3**, 367–382.
- Ito, J. (1986). *Syllable theory in prosodic phonology*. Ph.D. dissertation, University of Massachusetts.
- Ito, J. (1990). Prosodic minimality in Japanese. *Papers from the parasession on the syllable in phonetics and phonology, Chicago Linguistic Society*, 213–239.
- Kahn, D. (1980). Syllable-structure specifications in phonological rules. In M. Aronoff & M.-L. Kean (Eds.), *Juncture* (pp. 91–105). Saratoga, CA: Anma Libri.
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica*, **5**, (Suppl. 5), 1–54.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, **51**, 2018–2024.
- Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122–147). Cambridge, MA: MIT Press.
- MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, **12**, 271–296.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, **25**, 71–102.
- Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187–210). Hillsdale, NJ: Erlbaum.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, **10**, 29–63.
- McCarthy, J., & Prince, A. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, **8**, 209–284.

- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, **18**, 1–86.
- McQueen, J. M. *Segmentation of continuous speech using phonotactics*. Manuscript submitted for publication.
- McQueen, J. M., & Cutler, A. (1997). Cognitive processes in speech perception. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 566–585). Oxford: Blackwell.
- McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, **10**, 309–331.
- 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, and Cognition*, **20**, 621–638.
- Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, London*, **B295**, 333–352.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298–305.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714–719.
- Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, **22**, 1, 93–136.
- Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1209–1228.
- van Ooijen, B. (1994). *The processing of vowels and consonants*. Ph.D. dissertation, University of Leiden, The Netherlands.
- van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, **24**, 573–583.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, **32**, 258–278.
- Pallier, C., Sebastián-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllable structure of spoken words. *Journal of Memory and Language*, **32**, 373–389.
- Peretz, I., Lussier, I., & Béland, R. (1996). The roles of phonological and orthographic code in word stem completion. In T. Otake & A. Cutler (Eds.) *Phonological structure and language processing: Cross-linguistic studies* (pp. 217–226). Berlin: Mouton de Gruyter.
- Pinker, S. (1994). *The language instinct*. London: Penguin Books.
- Pitt, M. A., & McQueen, J. M. *Is compensation for coarticulation mediated by the lexicon?* Manuscript submitted for publication.
- Procter, P. (Ed.) (1975). *Longman dictionary of contemporary English*. London: Longman.
- Quené, H. (1989). *The influence of acoustic-phonetic word boundary markers on perceived word segmentation in Dutch*. Ph.D. dissertation, University of Utrecht, The Netherlands.
- Rissanen, J. (1989). *Stochastic complexity in statistical inquiry*. Teaneck, NJ: World Scientific Publishing.
- Schreuder, R., & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language*, **33**, 357–375.
- Sebastián-Gallés, N., Dupoux, E., Seguí, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language*, **31**, 18–32.
- Seguí, J. (1984). The syllable: A basic perceptual unit in speech processing? In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 165–181). Hillsdale, NJ: Erlbaum.

- Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24–49). Cambridge, MA: MIT Press.
- Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, 1239–1250.
- Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, **36**, 422–444.
- Swinney, D. A. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech* (pp. 201–209). Amsterdam: North-Holland.
- Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, **34**, 440–467.
- Thompson, H. (1990). Best-first enumeration of paths through a lattice: An active chart parsing solution. *Computer Speech and Language*, **4**, 263–274.
- Trager, G. L., & Bloch, B. (1941). The syllabic phonemes of English. *Language*, **17**, 233–246.
- Trubetzkoy, N. S. (1939). Grundzüge der Phonologie. *Travaux du Cercle Linguistique de Prague*, **7**.
- Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 98–108.
- Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, **24**, 744–755.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, **32**, 25–64.