Cues to speech segmentation: Evidence from juncture misperceptions and word spotting

JEAN VROOMEN and MONIQUE van ZON
Tilburg University, *Tilburg*, *The Netherlands*

and

BEATRICE de GELDER
Tilburg University, *Tilburg*, *The Netherlands*
and Université Libre de Bruxelles, *Brussels*, *Belgium*

The question of whether Dutch listeners rely on the rhythmic characteristics of their native language to segment speech was investigated in three experiments. In Experiment 1, listeners were induced to make missegmentations of continuous speech. The results showed that word boundaries were inserted before strong syllables and deleted before weak syllables. In Experiment 2, listeners were required to spot real CVC or CVCC words (C = consonant, V = vowel) embedded in bisyllabic nonsense strings. For CVCC words, fewer errors were made when the second syllable of the nonsense string was weak rather than strong, whereas for CVC words the effect was reversed. Experiment 3 ruled out an acoustic explanation for this effect. It is argued that these results are in line with an account in which both metrical segmentation and lexical competition play a role.

Understanding spoken language requires that listeners segment a spoken utterance into words or into some smaller unit from which the lexicon can be accessed. A major difficulty in speech segmentation is the fact that speakers do not provide stable acoustic cues to indicate boundaries between words or segments. At present, it is therefore unclear as to how to start a lexical access attempt in the absence of a reliable cue about where to start. Several decades of speech research have not yet led to a widely accepted solution for the speech segmentation problem. So far, three proposals have appeared in the literature that are of direct relevance here. One is that the continuous speech stream is categorized into discrete segments which then mediate between the acoustic signal and the lexicon. The second proposal is that there is an explicit mechanism that targets locations in the speech stream where word boundaries are likely to occur. The third is that word segmentation is a by-product of lexical competition. In the present study, these alternatives are considered.

Intermediating Units

One approach, which has been adopted by several psychological models of spoken word recognition, is to assume that the speech signal is classified into some intermediate prelexical linguistic unit. The notion is that the acoustic signal is categorized into segments, and once segments have been identified, lexical access can proceed without major difficulties. While there is, as yet, no agreement among psycholinguists about the structure or size of such a unit (e.g., phoneme, onset/rime, syllable, etc.), the syllable is clearly a segmentation unit that has captured attention. Several authors have claimed that speech is segmented into syllable-sized units (for an overview, see Segui, Dupoux, & Mehler, 1990). The basic idea of the "syllabic hypothesis" is that a lexical access attempt is initiated at the beginning of each syllable. A seminal study by Mehler, Dommergues, Frauenfelder, and Segui (1981) provided empirical evidence for such a syllable-based speech segmentation procedure. In their study, listeners detected a segment more quickly if it corresponded exactly to the first syllable of a word than if it comprised more or less than the syllable. Typically, listeners detected *ba* more quickly in *ba.lance* (the dot indicates the syllable boundary) than in *bal.con*, and *bal* more quickly in *bal.con* than in *ba.lance*. The benefit of syllable-based segmentation would be that the majority of lexical access attempts is successful, at least if contrasted with phoneme-based segmentation. However, an aspect that has put the syllabic hypothesis in a broader context is that linguistic variation appears to play an important role since perceptual procedures may depend on the listener's native language. The above-mentioned segment-detection results were obtained with French listeners and French stimuli. Subsequent studies showed that this pattern of results did not

hold up in English (Cutler, Mehler, Norris, & Segui, 1983, 1986). With English listeners, no syllabic effects were obtained; these listeners were equally fast in detecting *ba* or *bal* in *balance* and *ba* or *bal* in *balcony*. Cutler et al. (1986) attributed the asymmetric results to phonological differences between French and English. A major phonological contrast between these languages believed to be critical is the fact that English is a stress language with diverse syllable structures and English speakers' intuitions about syllable boundaries are often vague. In contrast, French has less diverse syllable structures and syllable boundaries are more clear. Cutler et al. (1986) argued that these factors made the syllable an appropriate segmentation unit for French but not for English.

Explicit Segmentation

The proposal made by Cutler et al. (1986) shifted attention from the now somewhat dated question about "the size of the intermediate unit" toward the issue of where in the speech signal word boundaries are likely to be perceived. At the same time, it introduced the notion that segmentation strategies of listeners were tuned to the phonology of the native language. The crucial aspect of the English phonology, and also the Dutch, is the metrical distinction between strong and weak syllables. Strong syllables have full unreduced vowels, whereas weak syllables have reduced vowels, which are usually realized as schwa. Words like *father*, *mother*, or *brother* all start with a strong syllable followed by a weak one, whereas words like *abuse*, *adjust*, or *believe* start with a weak syllable followed by a strong one. Cutler and Norris (1988) proposed the metrical segmentation strategy (MSS), which claims that English listeners initiate lexical access attempts at the beginning of every strong syllable. The speech recognition system thus takes the onset of strong syllables as the onset of lexical words (i.e., content words, excluding functors).

Prima facie evidence in favor of the MSS was obtained from the lexical statistics of the English vocabulary which, indeed, show that the success rate of the MSS will be quite high: Content words begin three times as often with strong syllables, and words beginning with strong syllables are twice as frequent as those beginning with weak syllables (Cutler & Carter, 1987). Words like *farther*, *mother*, or *brother* thus have a more typical stress pattern than words like *abuse*, *adjust*, or *believe*. Subsequent empirical evidence for the MSS came from two types of studies: juncture misperceptions and word spotting. Cutler and Butterfield (1992) examined mislocalizations of word boundaries in continuous speech. They presented sentence fragments to listeners at a level just above their threshold for speech perception. These barely audible sentences consisted of strings of alternating strong (S) and weak (W) or weak and strong syllables (e.g., *conduct ascents uphill,* which has a WS WS WS stress pattern; example taken from Cutler & Butterfield). Listeners showed a strong tendency to insert erroneous word boundaries before strong syllables and to delete word boundaries before weak syllables (e.g., *conduct ascents uphill → the doctor sends her bill* with a W SW S W S pat-

tern). Thus, in accordance with the MSS, listeners seemed to rely on a strategy of assuming that strong syllables marked the beginning of lexical words.

A second line of empirical evidence favoring the MSS came from a word-spotting study (Cutler & Norris, 1988). Listeners were required to monitor bisyllabic pseudowords and to press a button as soon as they heard a real word embedded at the beginning of such a pseudoword. The listeners monitored for CVC (e.g., *thin*) or CVCC (e.g., *mint*) words (C = consonant, V = vowel) that were embedded in a pseudoword string that ended in either a strong (e.g., *thintayf* or *mintayf*) or a weak syllable (e.g., *thintef* or *mintef*). In the case of a strong syllable (*thintayf* and *mintayf*), the MSS predicts that the pseudoword will be segmented as *thin_tayf* and *min_tayf* (the underscore indicates the metrical segmentation boundary), whereas there is no segmentation at all in the case of a weak syllable ending (*thintef* and *mintef*). In line with these predictions, the results showed that CVCC words like *mint* were harder to detect in *mintayf* than in *mintef*, whereas there was no difference for CVC words: *thin* embedded in *thintayf* was detected as quickly as *thin* embedded in *thintef*. It was proposed that the CVCC target *mint* from *mintayf* was divided across two segmentation units into *min_t*, with the impeding consequence that speech material had to be assembled across a segmentation boundary. For CVC words (*thin*) there was no difference between *thintayf* and *thintef* because the segmentation trigger in *thin_tayf* did not penetrate *thin*.

A Language-Universal Account:
Rhythmic Segmentation

The metrical effects observed in English and the seemingly different syllabic effects observed in French have recently been combined in an approach that covers the differences between these two languages. The more general proposal is that speech segmentation is based on language rhythm (Cutler, Mehler, Norris, & Segui, 1992; Cutler, Norris, & McQueen, in press). The rhythm of English can be characterized as stress-based, whereas French has syllabic rhythm. This argument is in line with studies showing that English listeners apparently use stress-based segmentation (Cutler et al., 1986) and French use syllabic segmentation (Mehler et al., 1981). Moreover, this more general proposal led to the prediction that moraic segmentation should be found in Japanese, which has moraic rhythm. And, indeed, this prediction was confirmed in a study showing that the mora was a relevant segmentation unit for Japanese listeners (Otake, Hatano, Cutler, & Mehler, 1993). The general notion is thus that phonological differences between languages are reflected in the segmentation procedures of their native listeners.

Lexical Competition as a Mechanism
for Speech Segmentation

The idea that segmentation strategies are adapted to the rhythmic structure of the native language may need to be extended in light of the more recent findings of Norris, McQueen, and Cutler (1995) and Vroomen and

de Gelder (1995). In Norris et al.'s study, the focus was on whether lexical competition played a role in speech segmentation. The concept of interword competition as a mechanism for speech segmentation is important in models like TRACE (McClelland & Elman, 1986) or Short-list (Norris, 1994), where segmentation emerges as a consequence of lexical competition. In TRACE, words inhibit each other to the extent that they overlap, and this inhibition serves as a segmentation device. Norris et al. (1995) investigated lexical competition effects using a word-spotting task in which subjects had to detect CVC or CVCC words with few or many competitors. Competitor size of the target words was defined as the number of words that have the second syllable of the nonsense string in which the target is embedded as onset. Thus, the competitor size of the target *mint* embedded in *mintayf* is equal to the number of words in the lexicon that start with *tayf*. They predicted that lexical competition would be larger for words with many competitors. Norris et al. replicated the MSS effect for CVCC words (i.e., *mint* easier to detect in *mintef* than in *mintayf*), but they also observed a competition effect for CVC words. When CVC words had many competitors, recognition was *facilitated* when compared with CVC words with few competitors. For example, the word *pram* embedded in *prampidge* was detected faster than *thin* embedded in *thintaup*, presumably because there are more words in English starting with *pidge* than with *taup*. In light of that evidence, the authors concluded that lexical competition and metrical segmentation might operate together.

The same conclusion was reached by Vroomen and de Gelder (1995), using a cross-modal repetition priming paradigm. In their study, the separate or combined effects of speech segmentation based on strong syllables and lexical competition were investigated. Subjects heard Dutch CVCC (e.g., *melk*, milk) or CVC words (e.g., *bel*, bell) embedded in bisyllabic nonsense strings. The second syllable was either weak (*melkem* and *belkem*) or strong, and the cohort size of competitors (as defined previously) starting with strong syllables was either small (*melkeum* and *belkeum*) or large (*melkaam* and *belkaam*—in Dutch, there are few words starting with *keum* and many starting with *kaam*). These auditory nonsense words served as prime for a visual target (MELK or BEL). In the CVCC words, where there is overlap between the embedded target and its competitors, it was observed that *melkem* had the largest facilitatory effect on MELK, *melkeum* had an intermediate effect, and *melkaam* had the smallest effect. For CVC words in which there is no overlap between the target and its competitors and thus also no competition, there was no difference in the facilitatory effects of *belkem*, *belkeum,* and *belkaam* on BEL. Priming effects of CVCC words, but not of CVC words, were thus proportionate to the number of competitors. These results were interpreted as the joint operation of metrical segmentation (because weak syllable endings do not activate a cohort of competitors) and lexical inhibition (because a small cohort of competitors has less of an impact on priming effects than does a large cohort).

The Present Study

So far, the rhythmic hypothesis has generated cross-linguistic comparisons between metrically different languages (i.e., French, English, and Japanese). In the most general terms, the finding is that different languages yield different results that are a function of the metrical characteristics of the language. These conclusions have often been reached on the basis of different paradigms such as fragment detection, word spotting, or priming which are, however, not always directly comparable to each other. For the language-universal claims of the rhythmic segmentation hypothesis, however, while it is important to look at differences between different languages with different tasks, it is equally important to find similarities between metrically similar languages using similar tasks. A critical issue that has so far not been addressed is whether metrically similar languages are covered by the language-universal rhythmic segmentation hypothesis as well. This may, in fact, turn out to be an even stronger test case for the rhythmic segmentation hypothesis, because different languages may have potentially important differences in phonology, distributions of lexical properties, and so on, which may all play a role. At present, it is unknown whether any of these nonmetrical characteristics are important for the results obtained so far. It is therefore of crucial importance to conduct studies in languages with comparable metrical characteristics so that the notion of rhythmic segmentation can be deconfounded. The present study is a step in this direction. Given that Dutch is stress-based, support for strong syllable segmentation would be support not only for the MSS, but also for the language-universal claims of rhythmic segmentation.

The critical question addressed here is whether a segmentation procedure that has been proposed for English, and that is based on the rhythmical properties of English, is also relevant to another language, one that has similar rhythmic properties. For phonological reasons similar to those given for English, Dutch seems to be a candidate for testing the applicability of the MSS. The lexical statistics of Dutch support the MSS inasmuch as an overwhelming majority (87.7%) of Dutch lexical words start with a strong syllable in initial position (see Vroomen & de Gelder, 1995). Moreover, Dutch, like English, has various syllable structures (up to CCCVCCC syllables, as in *strengst*; most strict), and many syllables have opaque syllable boundaries (e.g., *ba[ll]et* where the [ll] is an ambisyllabic consonant that belongs to both syllables). A syllabic segmentation routine as has been proposed for French (Mehler et al., 1981) is therefore not expected to apply in Dutch. On the other hand, syllabic effects in Dutch have been reported by Zwitserlood, Schriefers, Lahiri, and van Donselaar (1993). They observed that, as in French, segment-detection latencies were shorter if the target exactly matched the first syllable of a spoken word. This conclusion, however, could not be corroborated by Vroomen and de Gelder (1994), who also used a segment-detection task but different items. Similarly, Cutler (personal communication, 1995), using the original French items of Mehler et al. (1981), could not replicate, with

Dutch listeners, the syllabic effect reported by Zwitserlood et al. So the status of the syllable for speech perception in Dutch is unclear, and the present study might indirectly shed some light on this issue.

Given that lexical and phonological characteristics of Dutch are similar to those of English, the question is whether Dutch listeners actually apply an MSS-like strategy. To address this issue, three experiments were conducted. In the first, we used the juncture-misperception paradigm as introduced by Cutler and Butterfield (1992). Subjects were presented with barely audible strings of Dutch words made up of strong and weak syllables. Following the predictions of the MSS, one would expect that erroneous word boundaries would be inserted before strong syllables and deleted before weak syllables. One would also expect word-class effects. As in English, most lexical words start with strong syllables, but such unmarked grammatical words as *de* (*the*, masculine or feminine) or *het* (*the*, neuter) are usually realized with a single weak syllable. A word-initial strong syllable is thus most likely the onset of a lexical word, whereas a weak syllable is likely to be a grammatical word. One expects, therefore, that boundaries erroneously inserted before strong syllables produce lexical words, whereas boundaries inserted before weak syllables produce grammatical words.

The second experiment used the word-spotting task used by Cutler and Norris (1988). Subjects spotted words that corresponded to the initial CVCC (e.g., *melk*, milk) or CVC (e.g., *bel*, bell) fragment of a bisyllabic pseudoword. The second syllable of this pseudoword was metrically strong (i.e., containing a full vowel, as in *melkoos* or *belkoos*) or weak (the vowel was a schwa, as in *belkes* and *melkes*). Since there are very few words in Dutch that start with unvoiced plosives followed by a schwa, the number of competitors (as defined by Norris et al., 1995) for a target followed by a weak syllable is small, whereas the competitor size for targets followed by a strong syllable is large. If segmentation in Dutch is like that in English, responses for CVCC words followed by a strong syllable should be slower than those followed by a weak syllable (detection of *melk* slower in *mel_koos* than in *melkes*). For CVC words, one might expect a lexical competition effect as in Norris et al., such that detection of *bel* in *belkoos* is easier than detection of *bel* in *belkes* because there are many more words that start with *koos* than there are that start with *kes*. Finally, Experiment 3 served as a control experiment to check whether the observed effects could be explained by acoustic differences.

A possibility one should consider beforehand is that of syllabic segmentation. If it is true that, as suggested by Zwitserlood et al. (1993), Dutch listeners apply a syllabic strategy, one would expect that target words would be detected faster if they corresponded to the first syllable of the pseudoword. Most phonologists would agree that pseudowords such as *melkoos*, *belkoos*, *melkes*, and *belkes* are syllabified as *mel.koos*, *bel.koos*, *mel.kes*, and *bel.kes* (see, e.g., Collier & de Schutter, 1985). At first sight, then, Dutch CVC words should be detected faster than CVCC words, since the latter, though not the former, straddle a

syllable boundary. This comparison, however, is confounded in many ways. First, there are many (unknown) item differences between CVC and CVCC targets (among others, frequency of occurrence, length, phonetic make up, etc.) that may all play a role in word spotting. For these reasons, we refrain from making any direct comparisons between CVC and CVCC targets. Moreover, it is somewhat crude to contrast syllabic versus metrical effects as if they were two competing candidates. In fact, both may play a role just as acoustic, phonetic, or lexical effects do. The present study is therefore not intended to refute either the syllabic or a metrical hypothesis, as both may be applicable. Rather, the critical aspect is whether there is an independent contribution of metrical segmentation besides all other factors that are important. If so, one should find an effect of the strength of the second syllable in CVCC words. That is, if metrical segmentation is at stake, there should be a difference in detecting *melk* embedded in *melkes* versus *mel_koos*.

EXPERIMENT 1

Experiment 1 was similar to the laboratory-induced mis-segmentation experiment of Cutler and Butterfield (1992), this time using Dutch listeners and stimuli. The listeners heard barely audible sentence fragments which they had to report. Participants were expected to demonstrate word-boundary misperceptions, inserting erroneous word boundaries before strong syllables and deleting them before weak syllables; boundaries inserted before strong syllables should produce lexical words, boundaries inserted before weak syllables should produce grammatical words.

Method

Subjects. Twenty-one university students participated. They were all native speakers of Dutch, and none of them reported any hearing disorders. They were paid a small amount for participation.

Pretest materials and procedure. To estimate for each listener an individual speech-perception threshold, the procedures were similar to those of Cutler and Butterfield (1992). Two pretests were conducted for each participant. For the first pretest, a short passage of a newspaper text was recorded by a male speaker of Dutch. For the second pretest, 36 spondees (i.e., words with two strong syllables, such as *kaasboer*, cheese-maker) were recorded by the same speaker. All recordings were made in a studio. The materials were played in a soundproof booth over Sony MDR CD450 headphones from a Philips 850 DAT recorder connected to a step attenuator. The attenuator was calibrated with a 1-kHz signal. A Fluke 8922A decibel meter connected to the headphone indicated that one step on the attenuator was equal to approximately .25 dB.

Pretesting started with the passage from the newspaper played back at a comfortable listening level. The listener was asked to adjust the volume knob to the lowest level at which he could still understand the speaker. Some questions about the materials were asked at the end to confirm that participants had been able to follow the speech at the volume level they had chosen. This individually adjusted volume level served as the starting point for the second pretest, in which subjects were presented with the spondees, which they were asked to repeat. For each three correct consecutive repetitions, the volume was decreased by three steps on the attenuator until one word was repeated incorrectly. After an incorrect word, the volume on the attenuator was increased one step at a time until an item was repeated correctly. The level at which the participant re-

sponded 50% correct was, as in Cutler and Butterfield (1992), the level at which testing started.

Experimental materials. Fifty-four sequences of six syllables were constructed. A sequence consisted of monosyllabic or bisyllabic words with an unpredictable alternation of strong (S) and weak (W) syllables (e.g., the sentence *vroeger bracht gezang ons*—earlier brought singing us—has a stress/word boundary pattern as in SW S WS S). The word sequence was semantically unpredictable, but syntactically correct. To make them less predictable, the fragments were not complete sentences. In contrast to Cutler and Butterfield (1992), we did not use strictly alternating WS or SW sequences of strong and weak syllables. Rather, in the present case, the sequences of strong and weak syllables were more random, such that the stress pattern could be considered somewhat less predictable and more natural. Note that the stress pattern by itself can be divided in many different ways (e.g., SW S WS S can be segmented as S W S WS S, W SW SS, SWS WSS, etc.). There was thus ample opportunity in the material for word-boundary deletions or insertions to occur before weak or strong syllables. Ignoring the first syllable, since subjects have to assume that it is word initial, 67% ($n = 182$) of the syllables were strong and 33% ($n = 88$) were weak. Fifty-six percent of the strong syllables ($n = 102$) were word initial and 47% ($n = 42$) of the weak syllables were word initial (see Appendix A for the materials).

Design and Procedure. The sequences were recorded by the same speaker as in the pretest. The peak level of the strong syllables on the VU meter was approximately equal for each sequence. A sequence was repeated twice. Prior to each trial, the number of the trial was given, and prior to each repetition, the word "again" was recorded. Both the number and the word "again" were recorded several decibels above threshold.

Participants were tested individually. They were told that they were going to listen to speech presented "as if the radio was on a low volume." Their task was to write down what they thought had been said. They were asked to mark a dash if they were sure that a syllable had been spoken but were unable to report which one. This allowed us to analyze responses on which subjects had reproduced the correct number of syllables.

Results

The analysis of results was similar to that done by Cutler and Butterfield (1992). There was a total of 1,134 responses (21 subjects × 54 sequences), but only the responses that had (1) the same rhythmic pattern as the input and (2) the same number of syllables (six syllables) were analyzed. Since the goal was to analyze misperceptions, responses that were entirely correct (205) and responses with more or less than six syllables or with a different rhythmic pattern from that of the input (734) were discarded. The total number of responses that fulfilled the criteria was 195. Thus, 17% of all responses was analyzed, which is more or less similar to the 19% Cutler and Butterfield were able to analyze (i.e., 168 out of 864 responses). Within the 195 responses, 282 word-boundary misplacements were made, with several responses containing more than one word-boundary error (cf. Cutler & Butterfield, 1992, who obtained 264 word-boundary errors). There were 137 word-boundary insertions and 145 word-boundary deletions.

Table 1 presents some examples of the responses given. Examples of all four types of word-boundary misplacements occurred: insertions of a word boundary before strong syllables (e.g., *intern → in kern*, internal → in root), insertions before weak syllables (*minder → vindt het*, less → finds it), and word-boundary deletions occurred before strong syllables (*kreupel loopt → kreukeloos*, limpingly walks → wrinkleless), and before weak syllables (in*tern be*sluit → de *kerker* sluit, internal conclusion → the jail closes).

In the statistical analyses on these data, a goodness-of-fit measure was computed where the frequency of the expected number of word-boundary misplacements was compared with the observed frequencies. The expected frequencies were based on the actual properties of the stimulus input. We thus computed the number of weak and strong word-initial and non-word-initial syllables from the 195 sequences in which errors were made that fulfilled the criteria. The total number of syllables was 975 (195 sequences × 5 syllables, discarding the first syllable); 358 of these syllables (36.7%) were strong word-initial, 292 syllables (29.9%) were strong non-word-initial, 190 syllables (19.4%) were weak word-initial, and the remaining 135 syllables (13.8%) were weak non-word-

Table 1
Examples of Slips of the Ear

	Input	Error
Deletion before: weak	je eerder *zelf* beweerd "you earlier self asserted"	die eerder *zelfbe*heer "that earlier self-manage"
	in*tern be*sluit gezien "internal conclusion seen"	de *kerker* sluit gezien "the jail closes seen"
Deletion before: strong	uw leeftijd kreu*pel loopt* "your age limpingly walks"	in leeftijd kreu*keloos* "in age wrinkleless"
	de zieke eerder kramp "the patient earlier cramp"	*bezoeken* eerder dan "visit earlier than"
Insertion before: weak	de koffie *geurde* sterk "the coffee smelled strong"	de koffie *geurt te* sterk "the coffee smells too strong"
	je moeilijk *minder* geld "you difficult less money"	je moeder *vindt het* wel "your mother finds it surely"
Insertion before: strong	vroeger bracht *gezang* ons "earlier brought singing us"	vroeger bracht *de zang* ons "earlier brought the song us"
	beroemd *gedicht* gemaakt "famous poem made"	beroemd*ste vis* gemaakt "most-famous fish made"

initial. The expected number of errors corresponded to these input properties. That is, word-boundary deletions may occur before word-initial syllables (strong or weak) and word-boundary insertions may occur before non-word-initial syllables (strong or weak). For example, 36.7% of the input syllables were strong word-initial syllables. The expected chance of deleting a word boundary before such a word-initial strong syllable is therefore .367, which corresponds to 84.3 errors on the total of 282 word-boundary errors. The observed number of erroneous word-boundary insertions and deletions and the expected frequencies are presented in Table 2. As can be seen, in accordance with the predictions of the MSS, insertions before strong syllables and deletions before weak syllables occurred more often than they would by chance [$\chi^2(3) = 19.13, p < .001$].

We also compared the number of expected and observed frequencies for each individual subject. Of the 21 subjects, 14 produced, as predicted by the MSS, more insertions before strong syllables and more deletions before weak syllables, 2 subjects had the opposite pattern, and there were 5 ties. This number is significantly different from chance ($z = 1.83, p < .04$). Separately, by type of error, 15 subjects had more insertions before strong syllables, with 1 tie ($z = 2.01, p < .03$) and 18 subjects had more deletions before weak syllables ($z = 3.05, p < .005$).

Because we had repeated measures on the items, we could also perform an item analysis. The item analysis is, however, restricted because not every sequence had input characteristics that allowed all word-boundary errors to occur (insertions and deletions before strong and weak syllables). Moreover, there were several sequences in which no errors that fulfilled criteria were made. There were therefore a large number of ties in the item analysis. Nevertheless, of 54 sequences, 7 produced more insertions before strong syllables and more deletions before weak syllables, 3 had the opposite pattern, and the rest of the 44 sequences were ties ($z = .949, p = .17$). Separately, by type of error, 12 sequences had more insertions before strong syllables, 6 had the opposite pattern, and there were 36 ties ($z = 1.179, p = .12$). For deletions, 16 sequences had more deletions before weak syllables, 5 had the opposite pattern, and there were 33 ties ($z = 2.182, p < .02$).

Table 3 shows the distribution of the word classes with their expected frequencies after an erroneous boundary insertion (note that in this case expected frequencies are computed on the basis of the product of rows and columns because we do not have a basis for estimating the tendency to produce lexical or grammatical words). We excluded dashes and nonwords from the analyses. As predicted by the MSS, lexical words are more often produced when the erroneous word boundary precedes a strong syllable, whereas grammatical words are more often produced when the boundary precedes a weak syllable [with correction for continuity, $\chi^2(1) = 16.94, p < .001; z = 1.727, p < .05$, for lexical words; $z = 3.73, p < .001$, for grammatical words].

Discussion

In this first experiment, the pattern of word-boundary misplacements is found to be the same as it is for English and as predicted by the MSS: Listeners insert word boundaries before strong syllables and delete them before weak syllables; boundaries inserted before strong syllables tend to produce lexical words, boundaries inserted before weak syllables tend to produce grammatical words. Dutch listeners thus seem to treat strong syllables as the onset of lexical words, and weak syllables as non-word-initial; if word-initial, they are more likely to be grammatical words. This pattern of results closely corresponds to that obtained for English, and it thus confirms the claims of the rhythmic segmentation hypothesis. However, the empirical basis of the MSS hinges not only on juncture misperceptions; word-spotting data are equally important. The next two experiments therefore used the word-spotting paradigm introduced by Cutler and Norris (1988) to determine whether Dutch participants would employ an MSS in word spotting.

EXPERIMENT 2

Dutch listeners were required to spot real CVCC (e.g., *melk*, milk) or CVC words (e.g., *bel*, bell) embedded in bisyllabic pseudowords. The second syllable of the pseudoword was either weak (*melkes* or *belkes*) or strong (*melkoos* or *belkoos*). If listeners are guided by the MSS, one expects that a segmentation trigger is set at the onset of a strong syllable such that *melkoos* and *belkoos* are segmented as *mel_koos* and *bel_koos*. No segmentation trigger should be set for *melkes* and *belkes*. Detection of *melk* should therefore be harder in *mel_koos* than in *melkes*. If only the MSS is applied, there should be no difference in the detection of *bel* in *bel_koos* or *belkes*. But if lexical competition is at stake, as in Norris et al. (1995), one might expect that *bel* in *belkoos* would be easier to detect than *bel* in *belkes* because there are many more words starting with *koos* than there are with *kes*.

Method
Materials. Forty-two words were selected; half of them ended in a consonant cluster, and half ended in a single consonant. The final consonant of the cluster was always a stop consonant. As in Cutler and Norris (1988), the words formed pairs, such as *melk* (milk) and *bel* (bell), such that both words (1) had the same short vowel, (2) had the same postvocalic consonant, and (3) could not be made into words by adding or removing the second consonant from the coda (i.e., *mel* and *belk* do not exist in Dutch). All words were made into bisyllabic nonsense strings by the addition of an extra syllable. Two alternative VC endings were constructed: one had a strong

Table 2
Observed and Expected Word Boundary Insertions and Deletions Before Strong and Weak Syllables

		Observed	Expected
Insertions	Before strong	101	84.3
	Before weak	36	38.9
Deletions	Before strong	72	103.5
	Before weak	73	54.7

Table 3
Occurrence of Lexical and Grammatical Words and
Expected Frequencies Following Inserted Boundaries
Before Strong and Weak Syllables

	Before Strong		Before Weak	
	Occurrence	Expected	Occurrence	Expected
Lexical	51	44.5	12	16.5
Grammatical	3	9.5	10	3.5
Nonsense word or dash	47	—	16	—

vowel, the other was weak (schwa). The final consonant was constant within each pair. Thus, for the example given above, the endings were -oos/-es, making *melkoos, melkes, belkoos,* and *belkes*. The complete set of materials is presented in Appendix B. Analyses of the cohort sizes of the pseudoword endings showed that for the strong word endings there were an average of 334.6 words in the Dutch CELEX lexicon that start with the critical CV context as onset. Thus, for *melkoos* and *belkoos,* the critical context is *koo* and there are, on average, 334.6 words that have *koo* as onset. In contrast, words in the CELEX lexicon starting with an unvoiced consonant followed by a weak vowel in initial syllable position are very rare (in fact, there is one word starting with *ke,* five with *pe,* and five with *te*). Thus, for *melkes* or *belkes*, there are almost no words in Dutch that start with *ke* as onset. Words embedded in pseudowords with strong endings thus have many competitors; words embedded in pseudowords with weak ending have no or very few competitors. Another 80 bisyllabic nonsense strings were constructed that did not begin with a word. Forty of these strings ended in a full vowel; the other 40 ended in schwa. Examples are *wentoos, maspaat, wosper,* and *kalper*.

Two tapes were constructed, one for each version of each item. The type of context (SS, i.e., two strong syllables vs. SW, strong syllable first, second weak) was counterbalanced across word pairs and lists. Thus, *melkoos* and *belkes* appeared in one list, *melkes* and *belkoos* in the other. The nonsense strings were spoken in isolation by a male speaker of Dutch. The strings were digitized at 10 kHz, and then recorded on digital audio tape for presentation to subjects. All nonsense strings were spoken with primary stress on the first syllable. The interval between the trials was 3 sec. A short list of 16 practice trials was also recorded.

Subjects. Forty subjects were tested in a sound-attenuated booth. They were all students from the university and were paid a small amount. Half of them heard the first version of the stimulus set; the other half heard the second version.

Procedure. All subjects were tested individually. They were instructed that whenever they heard a nonsense string beginning with a real word, they should press the response key as quickly as possible and name the word they had detected into a microphone. The subjects' vocal responses were checked by the experimenter. Whenever a subject spoke any word other than the intended one, that response was discarded from subsequent analyses. The nonsense strings were presented over Sennheiser HD 410 SL headphones. A trigger aligned with the onset of the word started a reaction timer. Two reaction time (RT) analyses were made, one measuring RT from word onset, and the other, as in Cutler and Norris (1988), measuring RT from the onset of the burst of the embedded stop consonant. Thus, RTs for *belkes, belkoos, melkes,* and *melkoos* were adjusted by the length of the visually and auditorily determined onset of /k/. The mean adjustment length for CVC words was 305 msec in SS context and 280 msec in SW context; for CVCC words, the adjustments were 361 msec in SS context and 373 msec in SW context. Note that, for CVC targets, the adjustment amounts to the length of the embedded word. For CVCC targets, length is only partially compensated for as one should add the duration of the final consonant, which is, due to coarticulatory influences, difficult to determine. RTs of CVC and CVCC targets are thus confounded by different

length adjustments so that direct comparisons are difficult. However, as already argued, the difference between CVC and CVCC targets is not of interest in the present study as we were mainly interested in the effect of context.

Results

Responses to the items were inspected first. Two items (*park,* park, and *cent,* penny) were discarded from the analyses because, in later testing (see Experiment 3 for a full account), it appeared that the acoustic realization of the critical target word might have been different for one of the two tokens. For instance, when *park* was digitally excised from the SS-context *parkoes,* it was more difficult to recognize (missed by 71% of the listeners) than *park* excised from the SW-context *parkes* (in which case it was missed by only 5%). Similarly, *cent* excised from *centoos* was more difficult to recognize (miss rate of 47%) than *cent* excised from *centes* (miss rate of 5%). One of the reasons for these differences might have been that the acoustic realization of *park* in *parkoes* (or *cent* in *centoos*) was in a less canonical form than *park* in *parkes* (or *cent* in *centes*). Since we wanted to minimize acoustic artifacts, these items were excluded from subsequent analyses. To maintain the balanced structure of the item set, the matched CVC pairs were excluded as well. (It should be noted that removing these items was a conservative procedure since all items made a contribution in the predicted direction of the MSS.) This left 19 item quadruples on which subsequent analyses were based. Separate analyses of variance (ANOVAs) on RTs and error rates were conducted, with subjects and items as random factors.

Mean RTs and miss rates (i.e., no response to a target) for items and subjects were computed (Table 4). The RTs are measured from the burst onset of the stop consonant within the item. As can be seen, CVC words were detected somewhat faster than CVCC words, but this difference was not significant in the item analysis [$F_1(1,39) = 10.83$, $p < .002$; $F_2(1,18) = 1.33$, $p = .26$]. There was no difference in the latencies between the SS and SW context, nor was the interaction significant (in all cases, F_1 and $F_2 < 1$). Separate analyses for CVC and CVCC words on the RTs showed that the effect of context was not significant (in all cases, F_1 and $F_2 < 1$). Measuring RT from word onset did not change this pattern of results. RTs from word onset for CVC words were 1,213 msec in SS context and 1,225 msec in SW context; for CVCC words, the RTs were 1,281 msec in SS and 1,319 msec in SW context.

Analyses on the miss rates, however, present a different picture. Word spotting is a difficult task as many items are

Table 4
Mean Word Detection Times (in Milliseconds) and Miss Rates
for CVC and CVCC Items in SS and SW Context

	CVC			CVCC		
Context	Word	Detection Time	Miss Rate	Word	Detection Time	Miss Rate
SS	*bel*koos	828	.25	*melk*oos	920	.29
SW	*bel*kes	845	.34	*melk*es	946	.21

Note—S, strong syllable; W, weak syllable

missed. The overall miss rate in the present study was 27%, which is somewhat more than in Cutler and Norris's (1988) study, where the overall miss rate was 16% (Cutler, personal communication, 1994). The somewhat elevated miss rate in the present study might have been caused by the particular items that were selected (e.g., more low-frequency items), but such other factors as speaker characteristics or quality of the recording may also have played a role. Whatever the reason, the high miss rates justified an analysis on the number of misses. In the ANOVA on the miss rates, there was no main effect of context (both F_1 and $F_2 < 1$), and the main effect of target was significant only in the subject analyses [$F_1(1,39) = 4.75$, $p < .05$; $F_2 < 1$]. The important interaction between target type and context, however, was significant [$F_1(1,39) = 20.70$, $p < .001$; $F_2(1,18) = 10.58$, $p < .005$]. Separate analyses for CVC and CVCC targets showed that CVC targets were missed more often in the SW context than in the SS context [$F_1(1,39) = 16.18$, $p < .001$; $F_2(1,18) = 6.15$, $p < .03$]. Thus, a target word such as *bel* was more difficult to detect in *belkes* than in *belkoos*. The opposite was observed for CVCC targets: *melk* was more difficult to detect in the SS context *melkoos* than in the SW context *melkes* [$F_1(1,39) = 6.19$, $p < .02$; $F_2(1,18) = 5.43$, $p < .04$].

We also computed for each item the difference in miss rates for targets in the SS versus SW context. It is assumed that this difference is a somewhat purer measure of the influence of context on the target word, since idiosyncratic features of each item are in this way subtracted from each other. This difference score was correlated with the competitor size of the SS context. For CVC words, but not for CVCC words, the correlation was highly positive, indicating that the difference between targets from SS and SW contexts increased when the number of competitors in the SS context increased [$r(18) = .69$, $p < .001$]. The correlation thus indicates that CVC targets became easier to detect when followed by a string that was more likely to be the onset of a new word.

Discussion

The results of Experiment 2 show that CVCC words like *melk* are easier to detect in the SW context *melkes* than in the SS context *melkoos*. The opposite is the case for CVC words such as *bel*, which are easier to detect in the SS context *belkoos* than in the SW context *belkes*. The former finding partly replicates the results of Cutler and Norris (1988) in that a CVCC word such as *mint* was more difficult to detect in *mintayf* than in *mintef*. It should be noted, though, that the main difference in Cutler and Norris's study was in RTs rather than in error rates. However, although not reported in the original paper, the error rates in Cutler and Norris's study followed exactly the same pattern as in our experiment. That is, for CVCC words, error rates were 16.7 in the SS context *mintayf* versus 10.7 in the SW context *mintef*. For CVC words, the pattern was reversed: the error rate in the SS context *thintayf* was 16.7 versus 20.3 in the SW context *thintef* (Cutler, personal communication, 1994). Thus, in English too, there was a trend in that CVCC words were more difficult to spot in

the SS context than in the SW context, whereas the opposite was true for CVC words. Given that our data confirm this pattern, we conjecture that for CVCC words, the results are in line with the predictions of the MSS. The results for the CVC words, however, do not directly follow from the predictions of the MSS, but they are in accordance with a lexical competition account, as observed by Norris et al. (1995). In the framework of lexical competition, targets like *bel* in *belkoos* should be easier to detect than *bel* in *belkes* because *bel* in *belkoos* is followed by a string that is likely to be the onset of a new word. In contrast, *bel* in *belkes* is more difficult to detect because the *ke* string is not likely to be the onset of a new word. The correlation between the difference in SS and SW contexts and the competitor size showed that CVC words indeed became easier to detect when followed by a string that contained many words as onset.

At first sight, then, it seems that a combination of both the MSS and lexical competition can account for the present results. But before we elaborate on this interpretation, we need to examine the word-spotting data to determine whether they can be explained in acoustic terms. One might propose that CVC words are recognized better in the SS context because their acoustic realization is, in that case, in a more canonical form than it is in the SW context. It is, for instance, possible that there is more anticipatory assimilation of the final consonant of the CVC word in the SW context than in the SS context, and this coarticulation effect might have hampered recognition of the target word. To check for this possibility, another experiment was conducted in which the context was spliced from the target. As in Cutler and Norris (1988), we spliced, in the case of *melkes* and *melkoos,* the *es* and *oos* from the target-bearing pseudowords such that two *melk*s remained. Moreover, as we obtained an effect of context in CVC items, we also spliced the *kes* and *koos* from *belkes* and *belkoos* such that two *bel*s remained. If the nature of the context (strong or weak) is responsible for the observed pattern, splicing should have eliminated the difference between target words stemming from SS or SW context. There should then be no difference between *melk* taken from *melkoos* and *melk* taken from *melkes* or *bel* taken from *belkoos* and *bel* taken from *belkes*. On the other hand, if the observed pattern depends on the acoustic realization of the targets, splicing should have no effect on the observed pattern. In that case, should *melk* spliced from *melkes* be recognized better than *melk* spliced from *melkoos*, whereas *bel* spliced from *belkoos* should be recognized better than *bel* spliced from *belkes*?

EXPERIMENT 3

The third experiment was conducted to check whether the context or the acoustic realization of the target was the critical factor for the results obtained in Experiment 2.

Method

Materials. All experimental and nonexperimental items were made into monosyllables using a waveform editor. The final CVC

sequence was removed from the CVC words (*belkoos*, *belkes*) and the final VC was removed from the CVCC words (*melkoos*, *melkes*) so that *belkoos*, *belkes*, *melkoos*, and *melkes* became *bel*, *bel*, *melk*, and *melk*, respectively. For the fillers, the same procedure was applied: from half of them, the final CVC was removed so that they became CVC nonwords, and from the other half, the final VC was removed so that they turned into CVCC nonwords. For the CVC targets (*bel* from *belkoos* or *belkes*), splicing was done in the pause before the onset of the stop consonant of the second syllable. For the splicing of the CVCC targets (*melk* from *melkoos* or *melkes*), the splicing was done just before the first glottal pulse of the second vowel was visible so that as much as possible of the original item was included. As in the previous experiment, two tapes were made in which the spliced items appeared in exactly the same order as they had in the previous experiment.

Subjects. Forty subjects were tested in a sound-attenuated booth. They were all students from the university, and they were paid a small amount for participation. Twenty of them heard one of the two versions of the tape, and 20 heard the other version.

Procedure. The procedures were as close as possible to those of Experiment 2. Participants were asked to press a button whenever they heard a word, and then to say the word aloud. In the case of a nonword, no response was required. The vocal responses were checked by the experimenter.

Results

Preliminary analysis of the items showed that two tokens within an item pair differed markedly from each other. The target *park* excised from the SS-context *parkoes* was missed by 71% of the subjects, whereas *park*, excised from the SW-context *parkes*, was missed by only 5%. This is a 66% difference, which could, in principle, be accounted for by acoustic factors. Similarly, *cent* excised from the SS-context *centoos* was missed by 47% of the subjects, whereas *cent*, excised from the SW-context *centes*, was missed by only 5%. As we wanted to minimize the acoustic differences between the targets of the SS and SW contexts, we excluded these items from the analyses in the previous experiment and the present one as well. To maintain the balanced structure of the item set, we discarded the CVC matched item pairs. Similar analyses were then performed, as in Experiment 2. RTs were measured from word onset and from word offset. Mean RTs measured from word offset and miss rates for CVC and CVCC items are presented in Table 5.

In the ANOVAs on RTs, CVC words were detected somewhat faster than CVCC words, but this was significant only in the subject analysis [$F_1(1,39) = 13.05, p < .001; F_2 < 1$]. There was no difference between words excised from the SS or SW context, and the interaction between target type and context was not significant (all F_1 and $F_2 < 1$). Separate analyses for CVCC and CVC words showed that in none of these cases did the effect of context even approach significance (both F_1 and $F_2 < 1$). Measuring RT from word onset did not change this pattern. In this case, mean RTs were 779 and 788 msec for CVC words and 820 and 837 msec for CVCC words spliced from the SS and SW context, respectively.

Similar analyses were also performed on the miss rates. The results showed that there was absolutely no difference in the error rates between items excised from the SS or SW context (both F_1 and $F_2 < 1$). In the subject analysis, CVCC words were missed more often than CVC words [$F_1(1,39) = 33.04, p < .001$], but this difference was not significant in the item analysis [$F_2(1,19) = 2.35, p = .14$]. The important interaction between target type and context did not even approach significance (both F_1 and $F_2 < 1$). Separate analyses on the miss rates of CVCC and CVC words showed that in both cases the effect of context was not significant (all F_1 and $F_2 < 1$).

Discussion

In Experiment 3, CVCC items were somewhat more difficult to detect than CVC items, but this may be an artifact of the splicing procedure. One possibility is that a final stop consonant of a CVCC word is usually released, but when spoken in context, it is not. Due to the splicing procedure, the final consonant of CVCC items was unreleased, which made it sound somewhat unnatural. CVCC items might thus suffer more from splicing than would CVC items. The important result, however, is that the interaction between target type and context disappeared when the context was spliced from the target. Thus, *melk* spliced from *melkes* was as easy to detect as *melk* spliced from *melkoos*. The same pattern was also found for CVC words: *bel* spliced from *belkes* was as easy to detect as *bel* spliced from *belkoos*. This strongly suggests that the word-spotting results should be ascribed to the influence of the second syllable on the recognition of the target and not to the acoustic realization of the target itself.

GENERAL DISCUSSION

In the present study, we investigated whether speech segmentation was based on the language-specific rhythmic properties of a listener's native language. The claim of language-specific segmentation procedures cannot rest only on the observation of different segmentation procedures for phonologically contrasted languages. It is equally important to determine whether languages with similar phonological properties induce in their listeners similar segmentation procedures. The relevant aspect of Dutch is that it has a stress-based rhythm. This motivated

Table 5
Mean Word Detection Times (in Milliseconds) and Miss Rates
for CVC and CVCC Items Spliced From SS and SW Context

	CVC			CVCC		
Context	Word	Detection Time	Miss Rate	Word	Detection Time	Miss Rate
SS	*bel* from *belkoos*	407	.22	*melk* from *melkoos*	470	.31
SW	*bel* from *belkes*	422	.22	*melk* from *melkes*	470	.31

Note—S, strong syllable; W, weak syllable.

us to investigate whether the metrical segmentation strategy (MSS), as originally proposed by Cutler and Norris (1988) for English, was relevant for Dutch as well. The basic idea of the MSS is that listeners take strong syllables as the onset of lexical words. Finding evidence for strong syllable segmentation in Dutch would constitute evidence for the MSS beyond English, but more importantly, it would also confirm the claims of the language-universal rhythmic segmentation hypothesis.

In the first experiment, participants were induced to produce word-boundary errors while listening to speech fragments at a level just above threshold. As predicted by the MSS, word-boundary insertions were more likely to occur before strong syllables and word-boundary deletions were more likely to occur before weak syllables; word boundaries inserted before strong syllables tended to produce lexical words, and word boundaries inserted before weak syllables tended to produce grammatical words. These results correspond closely to those obtained for English listeners listening to English, and it thus seems that the MSS can account for the errors that occur when speech—Dutch or English—is hard to perceive.

In the following experiments, we used a word-spotting task to corroborate this conclusion. Subjects heard bisyllabic pseudowords and were asked to press a button as soon as they heard a real word embedded at the beginning of the nonsense string. The results showed that CVCC words were more accurately detected if followed by a weak syllable instead of a strong one: *melk* was easier to detect in *melkes* than in *melkoos*. This result is in line with the predictions of the MSS because a strong vowel should trigger segmentation of the CVCC word into CVC_C. Detection of *melk* in *melkoos* is thus difficult because the target is segmented as *mel_k*. However, an opposite pattern was observed for CVC words: *bel* was easier to detect in *belkoos* than in *belkes*. We have argued that the MSS on its own could not account for this result. At first sight, one might be tempted to argue that *belkoos* is segmented as *bel_koos* so that the segmentation trigger would make the end of the target more clearly marked if compared with *belkes*. There might thus be a benefit to be derived from the segmentation trigger if it correctly signals the *end* of the target word. However, it does not follow from the predictions of the MSS that a marked word ending should be of any help if compared with an unmarked ending: The MSS is about the *initiation* of a lexical access attempt, and not about the recognition process itself. Alternative explanations for these findings were therefore considered.

An intriguing possibility is that the word-spotting findings do not reflect only a metrical effect, but that they also result from lexical competition. In the TRACE model of spoken word recognition (McClelland & Elman, 1986) and in Shortlist (Norris, 1994), inhibition among lexical candidates depends on the number of phonemes that lexical items share within the same time slices. A CVCC word like *melk* in *melkoos* will be inhibited by words starting with *koo* or *koos* because these words are competing for /k/. There is thus competition at the lexical level for the

proper assignment of the acoustic input. As noted above, most lexical words in Dutch start with strong syllables, whereas there are no words that start with an unvoiced consonant followed by schwa. The targets from the SS conditions in the present study therefore had many competitors; targets in the SW condition had no competitors at all. Lexical competition for a CVCC word like *melk* in *melkoos* is therefore expected to be greater than that of *melk* in *melkes* because, in the former case, target and competitors are competing for the /k/. A word like *melk* in *melkoos* might therefore be more difficult to recognize because it is (1) more strongly inhibited via lexical inhibition than is *melk* in *melkes* and/or (2) because the metrical strategy sets a segmentation trigger in *mel_koos*. For CVC targets, the effects of lexical competition are different because there is no overlap between the target and its competitors. Nevertheless, it may be that a target like *bel* in *belkoos* is easier to detect than *bel* in *belkes,* because *koos* is more likely to be the onset of a new word than is *kes*. Thus, the chance of an erroneous assignment of the /k/ to the first word is lower in the *belkoos* case. *Bel* might therefore be easier to detect in *belkoos* than in *belkes* because its ending is more clearly marked.

It is, however, possible to see a complete picture of the intricate relations between metrical segmentation and lexical competition only if the results of different paradigms are compared. It is only through this comparison that it becomes clear when and how metrical segmentation and lexical competition emerge. It seems legitimate to argue that lexical competition and metrical segmentation selectively appear in quite different tasks and different circumstances, suggesting that both effects are independent of each other. Consider the case of CVCC items where there is overlap between target and competitor: In cross-modal repetition priming, it was observed that competitors *inhibit* the priming effect of CVCC targets but not of CVC targets (Vroomen & de Gelder, 1995). This contrasts with the word-spotting results. Here it seems that lexical competition has less impact on CVCC words inasmuch as we failed to observe a correlation between the number of competitors and the ease with which a CVCC target could be detected. Similarly, Norris et al. (1995), using word spotting, did not obtain a lexical competition effect in CVCC words. For CVCC targets, then, it appears that inhibitory lexical competition effects can be observed in cross-modal priming but not in word spotting.

The opposite pattern emerges for CVC items for which there is no overlap between target and competitor. In cross-modal priming, there was no effect of lexical competition on CVC targets (Vroomen & de Gelder, 1995), but in word spotting, competitors had a *facilitatory* effect. Thus, in the present study, we observed that CVC targets with many competitors were easier to detect than CVC targets with few competitors. Again, this result was also obtained by Norris et al. (1995) using English listeners. The question is how to account for these seemingly conflicting results. How is it possible that CVCC words, but not CVC words, suffer from competitors in cross-modal priming, whereas CVC words, but not CVCC

words, benefit from competitors in word spotting? One suggestion already alluded to may be that lexical competition has different effects, depending on whether or not there is overlap between target and competitor. A CVCC word such as *melk* in *melkoos* is competing with a cohort of *koos* words for the proper assignment of the critical phoneme /k/. This contrasts with a CVC word such as *bel* in *belkoos* which is not directly inhibited by words starting with *koo(s)*, because these competitors do not overlap with *bel*. This difference may help to explain why there is a difference in CVC and CVCC words across such tasks as word spotting and cross-modal priming. If one makes the assumption that cross-modal priming taps prelexical activation levels, competition effects may emerge early if competitors overlap with the target, thereby producing an inhibitory effect. These effects may disappear in the slower word-spotting responses, where they are masked by the much stronger metrical effects. On the other hand, the indirect competition effects for CVC targets may emerge only slowly over time. Since word-spotting responses are typically slow, this task may be sensitive to the indirect facilitatory competition effects, whereas responses in cross-modal priming may simply be too fast and already initiated before indirect competition could have its effects. It may thus be that the nature and the time course of the task determines whether facilitatory or inhibitory competition effects are observed. Inhibitory competition effects, which may arise early, can be found in a task that taps preactivation levels; facilitatory competition effects may arise late and can be observed in a task that taps recognition processes.

Taken together, the results from three different paradigms (cross-modal priming, missegmentations of continuous speech, and word spotting) strongly suggest the joint operation of lexical competition and metrical segmentation. Word-boundary errors produced by Dutch listeners can be accounted for by stress-based segmentation, whereas word-spotting data and cross-modal priming reflect metrical segmentation and lexical competition. As far as lexical competition is concerned, a determination needs to be made as to whether or not there is overlap between a target and its competitors. If there is overlap, inhibitory effects can be observed in cross-modal priming; if there is no overlap, facilitatory effects can be observed in word spotting. We favor this interpretation because there is now a growing amount of converging evidence from different paradigms and different languages (Cutler & Butterfield, 1992, using missegmentation; McQueen, Norris, & Cutler, 1994, and Norris et al., 1995, both using word spotting; Vroomen & de Gelder, 1995, using cross-modal repetition priming; the present study, using missegmentation and word spotting), all suggesting that metrical segmentation and lexical competition may give the speech-processing system a clue as to where word boundaries are likely to occur.

This proposal also raises important questions for future research: In contrast to rhythmic segmentation, lexical competition critically depends on the lexical properties of the language that can be distinguished from the rhythmic

characteristics. In contrast to rhythmic segmentation, lexical competition may be a more language-universal way to handle such peculiarities of the speech signal as the absence of word-boundary cues or the embedding of words in other words (see also de Gelder & Vroomen, 1994). At present, it still needs to be determined how *language-specific* segmentation procedures, that is, the mental processes that operate upon linguistic data, interact with *language-universal* procedures, such as interword competition, that operate on *language-specific lexical databases*.

REFERENCES

Collier, R., & de Schutter, G. (1985). Syllaben als klankgroepen in het Nederlands. *Antwerp Papers in Linguistics*, **47**.

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

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

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1983). A language specific comprehension strategy. *Nature*, **304**, 159-160.

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

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

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

Cutler, A., Norris, D., & McQueen, J. (in press). Lexical access in continuous speech: Language-specific realisations of a universal model. In T. Otake & A. Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies*. Berlin: Mouton de Gruyter.

de Gelder, B., & Vroomen, J. (1994). Metrical segmentation and lexical competition: A happy affair? *Dokkyo International Review*, **7**, 218-221.

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

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

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

Norris, D. (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, & Cognition*, **21**, 1209-1228.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory & Language*, **32**, 258-278.

Segui, J., Dupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification, and lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistics and computational perspectives* (pp. 263-280). Cambridge, MA: MIT Press.

Vroomen, J., & de Gelder, B. (1994). Speech segmentation in Dutch: No role for the syllable. In *Proceedings of the International Congress on Spoken Language Processing* (pp. 1135-1138). Yokohama, Japan.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 98-108.

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

APPENDIX A
Experimental Materials, Faint Speech

Sentence Fragment	Stress/Word Boundary Pattern	Sentence Fragment	Stress/Word Boundary Pattern
1. groot kasteel gewoond in	S SS WS S	28. een komisch leesboek ligt	W SS SS S
2. gebied als zee ontstaan	WS S S SS	29. moet protest in landen	S SS S SW
3. intern besluit gezien	SS WS WS	30. Chinees verzocht vergeefs	SS WS WS
4. 't leven buiten leidt	W SW W S	31. pastoor vertelt goedlachs	SS WS SS
5. arbeid zonder centen	SS SW SW	32. eerder niet gedacht te	SW S WS W
6. vies gebak met Nieuwjaar	S WS S SS	33. goed tehuis verzorgt de	S WS WS W
7. mooi verhaal verteld te	S WS WS W	34. uw leeftijd kreupel loopt	S SS SW S
8. de koffie geurde sterk	W SS SW S	35. je eerder zelf beweerd	W SW S WS
9. kiest bewust een heerschap	S WS W SS	36. kwamen vuisten onder	SW SW SW
10. forens bezocht volstrekt	SS WS SS	37. nieuwe buren komen	SW SW SW
11. onze eigen groente	SW SW SW	38. kontakt jaloers geweest	SS SS WS
12. karaf met goud versierd	SS S S WS	39. onder goud versta je	SW S WS W
13. was gejaagd en kattig	S WS S SS	40. hoort galant gedrag op	S SS WS S
14. Jan's student ontdekt het	S SS SS W	41. het eigen boek verkocht	W SW S WS
15. de zieke eerder kramp	W SW SW S	42. de lezing maandag stond	W SS SS S
16. verse kersen waren	SW SW SW	43. z'n prachtig rundvee kocht	W SW SS S
17. bekwaam beroep gehad	WS WS WS	44. dát moment verscheen hij	S SS WS S
18. miljoen of twee verkocht	SS S S WS	45. denken over Joden	SW SW SW
19. aan beide kanten kracht	S SW SW S	46. de moeder wees pardoes	W SW S SS
20. daar verwen je honden	S WS W SW	47. geschikt ballet bevat	WS SS WS
21. zíj goedkoop katoen in	S SS SS S	48. vroeger bracht gezang ons	SW S WS S
22. vijftig kikkers springen	SS SW SW	49. suiker had meteen in	SW S WS S
23. beroemd gedicht gemaakt	WS WS WS	50. géén verkleurd plafond in	S WS SS S
24. spion een goed motief	SS S S SS	51. je dolle zus verdacht	W SW S WS
25. gooide kluiten aarde	SW SW SW	52. je moeilijk minder geld	W SW SW S
26. díe pastoor noteert in	S SS SS S	53. goedkoop katoen gebreid	SS SS WS
27. neutraal en vaag herhaald	SS S S SS	54. naaide mooie weefsels	SW SW SW

APPENDIX B

CVCC (Freq)	SS, SW	CVC (Freq)	SS, SW
pont (4)	ponteus, pontes	non (19)	nonteus, nontes
*park (38)	parkoes, parkes	nar (1)	narkoes, narkes
link (2)	linkuut, linket	ding (371)	dinkuut, dinket
melk (51)	melkoos, melkes	bel (34)	belkoos, belkes
bink (1)	binkaar, binker	ring (34)	rinkaar, rinker
punt (172)	puntaal, puntel	dun (42)	duntaal, duntel
vamp (0)	vampool, vampel	ham (15)	hampool, ham-
pel			
vast (332)	vastoom, vastem	ras (25)	rastoom, rastem
mast (5)	mastoem, mastem	das (7)	dastoem, dastem
tulp (3)	tulpier, tulper	sul (1)	sulpier, sulper
kelt (1)	keltaaf, keltef	fel (61)	feltaaf, feltef
dank (79)	dankeet, danket	wang (67)	wankeet, wanket
hond (168)	hontuum, hontem	ton (30)	tontuum, tontem
*cent (26)	centoos, centes	den (7)	dentoos, dentes
milt (2)	miltoor, milter	pil (27)	piltoor, pilter
recht (232)	rechties, rechtes	pech (7)	pechties, pechtes
lift (28)	liftoos, liftes	rif (1)	riftoos, riftes
nest (24)	nestuum, nestem	zes (127)	zestuum, zestem
kalk (11)	kalkoom, kalkem	hal (30)	halkoom,
halkem			
mank (4)	mankoel, mankel	tang (5)	tankoel, tankel
hulp (116)	hulpoet, hulpet	nul (9)	nulpoet, nulpet
Mean = 61.8, SD = 91.5		Mean = 43.8, SD = 80.5	

*These quadruples were excluded from the analyses.