Chapter 2

Prosodic Structure and Word Recognition

Anne Cutler

2.1 Introduction

Prosodic structure is a dimension which belongs to spoken language. Although a good writer may aim for, say, rhythmic effects in prose, these rely upon the reader's ability to 'hear' them 'in the mind's ear', i.e. mentally to convert the written prose to a spoken form. As this chapter will outline, listeners make extensive and varied use of prosodic information in recognizing spoken utterances. However, because prosody is a property of spoken language, and because there has (purely for reasons of empirical tractability) been much less psycholinguistic research on spoken than on written language, the study of prosody's role in recognition is relatively underdeveloped. A recent comprehensive literature review in this area, covering the role of prosody in the comprehension of syntactic and discourse structure as well as in the recognition of spoken words (Cutler, Dahan & van Donselaar, 1997), lists some three hundred references, but this is a tiny amount compared with, for instance, the literature on visual word recognition, even that based on just one laboratory task (lexical decision). Moreover, as Cutler et al. conclude, the literature is very unbalanced: some topics have been repeatedly examined, in studies differing only in minor details, while other topics have been ignored completely. This is also true of research in different languages; as in all areas of Psycholinguistics, most research has been conducted in English, but among other languages some have received considerable research attention, some none at all. Particularly relevant here is the comparison between German and Dutch: the prosodic structure of these two languages is very similar, and has been comprehensively described for both languages in the phonetic literature, but although the psycholinguistic literature now contains a quite substantial number of experimental studies of the processing of Dutch prosody, there have been remarkably few comparable studies in German.

The present chapter concentrates on how prosodic structure can contribute to the recognition of words in spoken utterances. By prosodic structure is meant (as is generally assumed in phonetics and Psycholinguistics) the linguistic structure expressed in the suprasegmental properties of utterances. Note that there are other

would normally be spoken as weak syllables, with reduced vowels, therefore they cannot bear stress beats - *JACKson AND Jill WENT inLAND,* or *JACK and JILL climbed THE big HILL* are deeply unpleasant to the English ear.

There is abundant experimental evidence that English listeners make use of this stress-based rhythm to derive word-boundary information, namely by segmenting speech at the onset of strong syllables (i.e. those syllables with full vowels that can potentially be stressed). For example, when English-speakers make slips of the ear which involve misperception of word boundaries, they tend most often to insert boundaries before strong syllables (e.g., hearing *by loose analogy* as *by Luce and Allergy)* or delete boundaries before weak syllables (e.g., hearing *how big is it?* as *how bigoted?;* Cutler & Butterfield, 1992). Similarly, English listeners find word-spotting - detecting a real word embedded in a spoken nonsense sequence - hard if the word is spread over two strong syllables (e.g., *mint* in [m ntef]), but it is easier for them to detect a word spread over a strong and a following weak syllable (e.g., *mint* in [m nt f]; Cutler & Norris, 1988). Cutler and Norris argued that this difference arises because listeners divide [m ntef] at the onset of its second strong syllable, so that to detect the embedded word they must recombine speech material across a segmentation point; [m nt f], in contrast, offers no obstacles to embedded- word detection, as it is simply not divided, because the second syllable is weak.

Why should English listeners exploit stress rhythm in this way? Statistical studies of the English vocabulary and of distributional patterns in spontaneous speech (Cutler & Carter, 1987) have shown that a strategy of segmenting English at strong syllable onsets is in fact an extremely useful way of locating word onsets - most lexical words (nouns, verbs, adjectives) do indeed being with strong syllables. Distributional patterns in the Dutch vocabulary are similar to those of English - indeed, even more Dutch words than English words have a full vowel in the first syllable (van Heuven & Hagman, 1988; Schreuder & Baayen, 1994). Experiments modelled on those both of Cutler and Norris (1988) and Cutler and Butterfield (1992) have produced similar results in Dutch (Vroomen, van Zon & de Gelder, 1996). Thus the exploitation of stress rhythm seems to be a strategy which listeners use because it offers an efficient (partial) solution to the problem raised by the difficulty of location word boundaries in continuous speech. (German, like Dutch, has a very high proportion of words with full vowel in the first syllable [J. Bolte, personal communication], but relevant experimental evidence has not as yet been collected.)

Such a solution is, however, not open to speakers of languages without stress rhythm. French, for example, does not exhibit the type of contrast between strong and weak syllables observed in the Germanic languages. Experimental studies of the processing of spoken French suggest that listeners can draw on a process of segmentation of the input into syllable-sized units; thus in the words *palace* and *palmier* (which begin with the same three phonemes) the first two phonemes func-

tion as the initial unit in *pa-lace,* the first three in *pal-mier.* A wide variety of experimental tasks, involving prelexical processing, lexical processing, and representation of words in memory, produce results showing how important this procedure is in the recognition of French (Mehler, Dommergues, Frauenfelder & Segui, 1981; Segui. Frauenfelder & Mehler, 1981; Cutler, Mehler, Norris & Segui, 1986; Dupoux & Mehler, 1990; Kolinsky, Morais & Cluytens, 1995: Pallier, Sebastian-Galles, Felguera, Christophe & Mehler, 1993; Peretz, Lussier & Beland, 1996).

Although the use of stress-based rhythm in English and the use of syllabic segmentation in French might seem to be quite different solutions to the segmentation problem in continuous speech, they can also be viewed as similar: like stress in English, the syllable in French is the basis of rhythmic structure. This symmetry prompted the hypothesis (see, e.g., Cutler, Mehler, Norris & Segui, 1992) that listeners might in fact adopt a universally applicable solution to the segmentation problem, in that to solve it they exploit whatever rhythmic structure happens to characterize their language. This implies that if a language has a rhythmic structure based on some phonological construct other than stress or the syllable, it should be possible to find evidence for exploitation of such rhythmic structure in speech segmentation. Japanese is such a language; its rhythm is described in terms of a sub-syllabic unit, the mora (e.g., the word *tanshi* has three morae: *ta-n-shi).* Otake, Hatano, Cutler and Mehler (1993), Cutler and Otake (1994), and Otake. Hatano and Yoneyama (1996) conducted studies of prelexical processing by Japanese listeners, and indeed found consistent evidence favoring mora-based segmentation.

It is not the rhythmic structure of the input itself which produces the appropriate segmentation procedure; if this were so, then any listener could listen to any language and effectively 'hear' the word boundaries. Experience tells us that this certainly does not happen. Instead, it appears that listeners have developed segmentation procedures on the basis of experience with their native language, and that they do not command the appropriate procedures for other languages. Thus English listeners show no evidence of syllabic segmentation with French input (Cutler et al., 1986). and neither do Japanese listeners (Otake et al., 1996); English listeners likewise show no evidence of mora-based segmentation of Japanese input (Otake et al.. 1993; Cutler & Otake. 1994). and nor do French listeners (Otake et al., 1993) or Dutch listeners (Kakehi, Kato & Kashino. 1996). Moreover, listeners may apply their native language-specific procedures to foreign language input, even in cases where the procedures may not operate efficiently at all. Thus French listeners apply syllabic segmentation to English words such as *palace* and *palpitate* where English listeners do not (Cutler et al., 1986); likewise, they apply syllabic segmentation to Japanese input (e.g., preferring to segment *tanshi* as *tan-shi:* Otake et al., 1993); and Japanese listeners apply moraic segmentation where possible to English input (e.g., showing facilitated processing of the syllable-final nasal in words like *incur,* where English listeners do not; Cutler & Otake, 1994) and to French and Spanish input (e.g., re-

sponding equally rapidly to a consonant-vowel target such as *pa-* in an open and in a closed syllable; Otake et al., 1996).

Finally, it should be pointed out that the exploitation of different levels of linguistic structure in segmentation is not determined by the simple availability of these types of structure in a language. Every concept which has proved relevant in describing these cross-linguistic differences in segmentation - stress, syllable, mora - is a phonological construct which can in principle be applied to any language. It is the role of the relevant units in the rhythm of the language via which the units attain a role in segmentation. And even though the concept stress, for instance, is dependent on the concept syllable (stressed/unstressed are properties of syllables, not of parts of syllables), this does not accord the syllable the same role in a stress-rhythm language as it has in a syllable-rhythm language. Cutler et al. (1986) failure to find syllabic segmentation by English listeners has been replicated many times (Bradley, Sanchez-Casas & Garcia-Albea, 1993; Cutler, Norris & Williams, 1987; Kearns, 1994). In German, a study by Hohle and Schriefers (1995) found response patterns consistent with syllabic segmentation only for finally-stressed words with open initial syllables (i.e. *ku-* was detected more rapidly that *kul-* in *Kulanz).* In Dutch, Zwitserlood, Schriefers, Lahiri and van Donselaar (1993) found evidence of syllabic segmentation, but a comparable study by Vroomen and de Gelder (1994) found no such effects. Cutler (1997) presented Dutch listeners with the easily-syllabified French materials of Mehler et al. (1981); like the English listeners tested by Cutler et al. (1986), they did not show the response pattern shown by French listeners, although like Hohle and Schriefers' German subjects, they did respond faster to the syllabic targets in words with an open initial syllable (i.e. *pa-* was detected more rapidly than *pal-* in *palace).* Note that in the citation pronunciation of French words, accent falls on the final syllable, so that in both these cases the result is consistent with segmentation at the onset of a stressed syllable, the default segmentation in stress-rhythm languages.

Note that the stress rhythm of English, Dutch or German is not itself determined by word-boundary location; stress in these languages can occur at differing positions in the word. But in some languages stress is fixed, i.e. it must always fall at the same word-internal position. Fixed-stress languages include, for example, Finnish, in which the first syllable of every word is stressed, or Polish, in which stress always falls on the penultimate syllable. It might be imagined that fixed stress could provide an excellent cue to word-boundary location; but in fact, rather paradoxically, it is possible that fewer explicit acoustic correlates of stress may be available for listeners' use in fixed-stress languages than in free-stress languages. This is because the explicit realization of stress may be unnecessary when its location is fully predictable. Suomi. McQueen and Cutler (1997) carried out a segmentation experiment in Finnish, using the same word-spotting task as in the experiment by Cutler and Norris (1988) described above; the focus of their study was in fact not stress but

vowel harmony (which in Finnish requires that two vowels in the same word must be drawn from compatible classes). The listeners in Suomi et al.'s study heard bisyllabic words (e.g., *palo)* with preceding or following CV contexts *(kupalo, paloku);* all of the resulting trisyllabic nonsense items were spoken with the unmarked prosodic pattern for trisyllabic words, traditionally described as stress on the initial syllable. The principal result of Suomi et al.'s study was that vowel harmony functioned as a segmentation cue: words preceded by disharmonious contexts (syllables containing a vowel from a class incompatible with the vowels of the word) were detected more rapidly than words preceded by harmonious contexts. Stress was relevant for the interpretation of a control experiment, however, in which the words were excised from their contexts and presented to listeners in a lexical decision task; for the words from which preceding contexts had been removed, this resulted in loss of the syllable which had nominally been stressed, and these words might therefore have been expected to be prosodically abnormal compared to those from following contexts. Listeners' responses showed no effect attributable to abnormality of this kind, however; if anything, words like *palo* from *kupalo* were recognized slightly faster than words like *palo* from *paloku.* Suomi et al. suggested that the so-called initial stress of Finnish is actually a gradual drop in fundamental frequency and amplitude across the word, and that what is important for its correct realization is simply the relationship between consecutive syllables; this relationship would be unaffected by removal of preceding or following syllables.

French is another language with a consistent prosodic pattern which could provide information about some word boundaries. French does not have English-like stress, but accent falls on the final syllable of rhythmic groups, and the right boundary of a rhythmic group is always also the right boundary of a word. French listeners appear to be able to use this regularity to speed detection of a target syllable located at a rhythmic group boundary in comparison to the same syllable at another location (Dahan. 1996); the rhythmic structure thus indirectly facilitates lexical processing..

Prosodic structure, in the form of language rhythm, thus helps listeners in a number of ways to perform lexical segmentation efficiently. The characteristic rhythm of a language is undoubtedly real; it plays a role not only in lexical segmentation and other forms of processing, but most obviously in preferred poetic metres. However, it does not provide direct signals of word boundary location, but rather assists segmentation indirectly, establishing a framework within which listeners can hypothesize probable word boundary locations, or allowing lexical segmentation to proceed by inference from segmentation into higher-level constituents.

Early investigations of speech rhythm often assumed that rhythm should be directly encoded as regularity of timing of units in the speech signal; this line of research ended in rejection of the direct-encoding assumption by phonetic researchers (see Cutler, 1991, for a review). The possibility that rhythmic regularity existed and could be exploited by listeners was also addressed in psycholinguistic studies. Thus

Shields, McHugh and Martin (1974) presented listeners in a phoneme-detection study with nonsense words embedded in real sentences, and found that listeners detected the initial phoneme of the nonsense word more rapidly when the first (target-bearing) syllable was stressed rather than unstressed. However, the effect disappeared when the nonsense word was embedded in a string of other nonsense words, suggesting that the facilitation was not due simply to acoustic advantage. The authors concluded that the timing of speech events is predictable from temporal redundancy in the signal, and listeners can use the temporal structure to predict upcoming stresses.

Other studies supported this predictive view, by showing that disrupting the temporal structure impairs performance on many perceptual tasks. Martin (1979), for example, found that either lengthening or shortening a single vowel could cause a perceptible momentary alteration in the tempo of a spoken sentence, and increase listeners' phoneme-detection response times. Meltzer, Martin, Mills, Imhoff and Zohar (1976) found that phoneme targets which were slightly displaced from their position in normal speech (by deleting a short portion of the signal immediately prior to the target phoneme) were detected more slowly. Buxton (1983) found that when the word which preceded a target-bearing word in phoneme-detection was replaced by a different word, the replacement increased detection time to a greater extent if the two words differed in number of syllables. All these results were consistent with the proposal that listeners process a regular rhythm, using it to make predictions about temporal patterns; when manipulations of the speech signal cause these predictions to be proven wrong, recognition is momentarily disrupted.

Later results, however, called this interpretation into question. Mens and Povel (1986) conducted (in Dutch) an experiment modelled on that of Buxton (1983) in English, in which temporal modification was again achieved by replacing the pretarget word by one with a different number of syllables (e.g., *kat* - cat - was replaced by *kandidaat* - candidate). Mens and Povel failed to replicate the effects of predictability observed in the earlier studies. Pitt and Samuel (1990) similarly only weakly replicated Shields et al.'s (1974) result, in a phoneme-detection study using acoustically controlled target-bearing words embedded in natural sentence context. They found that predictability was only possible when the word was embedded in a rhythmically highly regular word list. Pitt and Samuel speculated that natural sentence contexts may in fact offer little opportunity for exercising prediction with respect to the location of stressed syllables.

A similar conclusion was also reached by Mehta and Cutler (1988), who found differences in the pattern of effects observed with spontaneously spoken versus read materials in a phoneme-detection experiment. One difference was that in read materials, but not in spontaneously spoken materials, targets occurring later in a sentence were detected faster than targets occurring earlier in a sentence; since the two sets of materials were identical in content (the former being read, by the original speaker, from a transcript of the spontaneous conversation), Mehta and Cutler concluded that

the effect in the read speech reflected not semantic predictability but temporal regularity. Together these results however suggest very little role for rhythmic regularity. To achieve real predictability which listeners can exploit at the word-by-word level, there must be sustained regularity, as Pitt and Samuel showed, and this appears not to occur at all in spontaneous speech; in read speech, it can come into play in longer sentences, in which the latter part of the sentences become easier to process than the earlier parts.

Using a word-monitoring task (in which listeners respond when they detect a specified target word), Tyler and Warren (1987) explored the effects of disrupting the temporal structure of meaningless (but syntactically acceptable) sentences as a function of the effects of this disruption on prosodic grouping; longer detection latencies were observed when phonological phrasing was disrupted, suggesting that grouping effects as realized in the prosodic structure play a stronger role than simple temporal predictability arising from regularity of rhythm. Rhythmic structure is used in lexical segmentation, but indirectly, in that it guides hypotheses about word-boundary location; however, rhythmic structure usually does not guide speech processing by allowing advance prediction of the speech structure itself.

However, there is one way in which the timing of speech events can be of assistance in word boundary perception - though this occurs not at the level of sentence prosody, but at the segmental level. Segmental timing varies with position in the word, and listeners can make use of this variation in segmentation. Thus Quene (1992, 1993) investigated minimal junctural pairs such as *naam op - na mop* in Dutch; he found that the lengthened duration of a consonant (especially sonorant consonants such as [m]) in word-final position was especially helpful to listeners. Overall syllable duration (but principally: duration of the vowel) formed a reliable cue in other studies: thus Nakatani and Schaffer (1978) found that relative syllable duration allowed listeners to distinguish English adjective-noun sequences such as *noisy dog* and *bold design* when these were presented as reiterant speech (in which a natural utterance is mimicked in a series of repetitions of a single syllable such as *ma),* and Rietveld (1980) reported similar results for French ambiguous strings (e.g., *le couplet complet - le couple est complet).* Such ambiguous strings may, of course, not often occur in natural speech; nevertheless temporarily ambiguous sequences do occur. Embedded words provide a case in point (thus *Stau* is embedded in *Staub,* which in turn is embedded in *Staupe);* listener sensitivity to segmental duration helps to avoid temporary ambiguity resulting from such embedding. Christophe, Dupoux, Bertoncini and Mehler (1994) showed that newborn infants can discriminate between Disyllabic sequences such as *mati* taken from within a word *(mathematicien)* versus across two words *(panorama typique);* relative syllable duration differed significantly across the two types of bisyllable. It is clear that human listeners do have finely tuned temporal-discrimination capacities, and these assist in segmentation just as sentence-level rhythm and grouping does.

Prominence

Words which bear sentence accent are processed more rapidly than words which do not. Thus targets on accented words are detected more rapidly than targets on unaccented word in phoneme detection (Cutler & Foss, 1977; Mehta & Cutler, 1988); verification of heard words is faster if the words were accented than if they were not (van Donselaar & Lentz, 1994); and mispronunciations are registered more rapidly in accented than in unaccented words (Cole, Jakimik & Cooper, 1978; Cole & Jakimik, 1980). Accented words have heightened acoustic clarity (increased spectral definition: Koopmans-van Beinum & van Bergem, 1989; and increased duration: Klatt, 1976, van Santen & Olive, 1990; Eefting, 1991; Dahan & Bernard, 1996), and this certainly could help to make them easier to process.

Nonetheless, the processing advantage of accented words is not solely due to acoustic factors. This is shown by Cutler's (1976) finding that the preceding prosodic contour leading up to an accented word in itself produces speeded processing. Cutler recorded sentences in two prosodic versions, one in which the word bearing the phoneme target was contrastively accented, and one in which contrastive accent fell elsewhere; for example, with the target phoneme *Id/: She managed to remove the DIRT from the rug, but not the berry stains; She managed to remove the dirt from the RUG, but not from their clothes.* The target-bearing word itself (i.e. in this case, the word *dirt)* was then edited out of each version and replaced by acoustically identical copies of the same word taken from a third recording of the same sentence, in which no contrastive accents had been applied. This resulted in two versions of each experimental sentence, with acoustically identical target-bearing words but with different prosodic contours on the words preceding the target: in one case the prosody was consistent with sentence accent occurring at the location of the target, in the other case it was consistent with accent falling elsewhere. Cutler found that subjects nevertheless responded significantly faster to the target in the 'accented' position than to the target in the unaccented' position. Since there were no acoustic differences between the target words themselves that could account for this result, and the only difference in the preceding context lay in the prosody, listeners must have been using this preceding prosodic information to predict where accent would occur.

A later study by Cutler and Darwin (1981) showed that this predicted accent effect was unaffected by the removal of pitch variation, i.e. the presentation of the sentences in a monotonized form; and it was also not affected by manipulation of the duration of closure for the target stop consonant. Thus the effect does not appear to be dependent on any particular prosodic dimension. When speech hybridization techniques were used to interchange timing patterns between the two versions of an utterance, however, so that 'impossible' utterances resulted (e.g., an utterance in which the FO contour suggested that accent would fall on the target-bearing word while the durational patterns of the preceding words suggested that it would fall elsewhere),

the predicted accent effect disappeared (Cutler, 1987), suggesting that consistency among the separate prosodic dimensions is important for listeners to be able to exploit them efficiently.

The effects of predictability which are robustly observed in these accent studies contrast with the fragile predictability effects observed when rhythmic regularity was at issue. Interestingly, in the experiments of Shields et al. (1974) and Meltzer et al. (1976), the target-bearing nonsense words seem to have been carrying the main information of the clause in which they occurred, so that the speaker would presumably have assigned them sentence accent. It is possible, therefore, that these authors unwittingly manipulated sentence accent as well as rhythmic structure, and that the effects that they observed were due to the former rather than the latter factor. The difference between the results of Buxton (1983) and Mens and Povel (1986) could have a similar root. Buxton's target-bearing words were nouns, and the rhythmic manipulation was carried out on an immediately preceding adjective; nouns are more likely to bear sentence accent than adjectives, so that it is likely that the manipulation in Buxton's study disrupted the prosodic structure immediately preceding an accented word. Mens and Povel, on the other hand, manipulated nouns and (in a minority of cases) verbs, and the following target-bearing word was in nearly all cases a preposition or adverb; their manipulation therefore most likely involved an accented word, while the target occurred in post-nuclear position in both the intact and the cross-spliced sentences.

Certainly the accent prediction effects do not appear to be based on any relation of sentence accent to the temporal structure of an utterance. Instead, it has been argued that listeners direct attention to accented words because these are semantically focussed, and hence convey information that is particularly important for apprehension of the speaker's message. Semantic focussing by itself leads to faster responses in phoneme detection in just the same way that prosodic accentuation does; Cutler and Fodor (1979) demonstrated this in a study in which semantic focus was manipulated by means of a question preceding the sentence in which the target occurred. Once located, focussed words receive more detailed semantic processing: Multiple meanings of homophones are activated if the words are in focus, but not necessarily if the words are not in focus (Blutner & Sommer, 1988), and recall of the surface form of a word is more likely if the word was in focus in a heard sentence than if it was not (Birch & Garnsey, 1995). Thus listeners may actively search for accented words because these provide the semantically most central portion of a speaker's message. Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard and Carlson (1995) demonstrated just how rapidly accent can be processed, in a study in which they tracked the eye movements of listeners who were required to select one of four items in a display. When the display set consisted of, for instance, a large red square, a large blue circle, a small red square and a small yellow triangle, and the listeners heard *touch the LARGE red square*, they were able to select the correct item on hearing the

contrastively accented word *large*. Apparently the contrastive accent allowed them to choose the one member of the set of large items which contrasted, by being large, with some other item.

The relation of accentuation to semantic structure is also underlined by a number of studies which show that listeners prefer, or find it easier to process, sentences in which accent falls on information which is new, i.e. has not previously occurred in the discourse (Bock & Mazzella, 1983; Terken & Nooteboom, 1987; Birch & Clifton, 1995). Syntactic disambiguation can also be effected by accent placement; Read, Kraak and Boves (1980) found that in the ambiguous Dutch sentence *me zoent de vrouw?* accent on the verb led listeners to prefer the interpretation, in which the woman is the subject and the object of the action is questioned. However, their explanation of this result drew on the relationship of accent to information structure: the accent on the verb effectively deaccented the following noun *(vrouw),* implying that it should be taken as existing topic of the discourse, which in turn implies that it is the grammatical subject of the sentence, and the question word is therefore the grammatical object.

The relative prominence of words which is conveyed by sentence prosody is thus exploited by listeners to derive information about the semantic relations within utterances; words which are accented effectively receive favored processing.

2.3 Word Prosody

Words may be uniquely distinguished from one another by differences in segmental structure (e.g., *Bein* from *mein, Bahn,* and *Beil);* but in many languages they may also be distinguished solely by suprasegmental means: *iibersetzen* from *ubersetzen*, in German (a stress language), *ame* with HL pitch accent (meaning *rain)* from *ame* with LH pitch accent (meaning *candy)* in Japanese, [si] with high level tone 1 (meaning *poem),* from [si] with high rising tone 2 (meaning *history),* from [si] with low level tone 6 (meaning *time)* in Cantonese. Thus word recognition in spoken-language understanding involves the processing of prosodic structure which may contribute to or even solely determine word identity.

The process of lexical access in spoken-word recognition is described in detail in this volume in the chapter by Frauenfelder. Current models of word recognition assume that multiple lexical candidates are activated by incoming speech input, and compete among one another for recognition. Both matching and mismatching information in the signal may contribute to a candidate word's fate: information in the signal which matches the stored lexical representation can increase the corresponding word's activation, while activation can be decreased by incoming information which fails to match what is stored. No current model of spoken-word recognition, whether computationally implemented or not, has as yet addressed specifically the role of prosodic information in this process. However, there is a large amount of

relevant experimental evidence available, which can shed light on such questions as whether prosodic information constrains the initial stages of lexical activation, or whether it plays a subordinate role by only coming into play in order to allow selection among alternative candidate words.

Lexical Tone

In lexical tone languages, words may be distinguished by the pitch height or contour of syllables, as in the Cantonese example given above. Thus only the single suprasegmental dimension of fundamental frequency (FO) is involved in signalling tone. This FO information can be highly informative even in the absence of segmental information; thus Ching (1985,1988) found that identification scores for lip-read Cantonese words improved greatly when FO information was provided, in the form of pulses synchronized with the talker's pitch (there was however very little improvement when FO information was provided for lip-read English words). Lexical priming studies in Cantonese suggest that the role of a syllable's tone in word recognition is analogous to the role of the vowel (Chen & Cutler, 1997; Cutler & Chen, 1995); in an auditory lexical decision task, overlap between a prime word and the target word in tone or in vowel exercise analogous effects.

Although it might seem that a contrast realized in FO should be perceptually simple to process (it resumbles, for instance, the contrast between two musical notes), listeners without experience with a tone language find tone discrimination difficult. Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda and Keller (1996) compared same-different discrimination of Thai tones and musical transformations of the same tones, by speakers of Thai, Cantonese and English; Thai and Cantonese listeners could discriminate the speech and musical tones equally well. but English listeners discriminated the musical tones significantly better than the speech tones. Lee, Vakoch and Wurm (1996) also found that English listeners had difficulty making same-different judgments on Cantonese or Mandarin tone pairs; speakers of the two tone languages always performed better than the English listeners (although they also performed better with the tone contrasts of their own language than with those of the other language).

Fox and Unkefer (1985) conducted one of the first psycholinguistic investigations of tone in word recognition, in a categorization experiment using a continuum varying from one tone of Mandarin to another. The crossover point at which listeners in their experiment switched from reporting one tone to reporting the other shifted as a function of whether the CV syllable upon which the tone was realized formed a real word when combined only with one tone or only with the other tone (in comparison to control conditions in which both tones, or neither tone, formed a real word in combination with the CV). This lexical effect appeared only with native-speaker listeners; English listeners showed no such shift, and on the control continua the two subject groups did not differ. Because the categorization task is not an 'on-line' task

(i.e. it does not tap directly into the process of word recognition), however, Fox and Unkefer's finding does not shed light on the issue of whether tone plays a role in initial activation of word candidates or only in selection between them.

However, tonal information may constrain word recognition less surely than segmental information. In a study by Tsang and Hoosain (1979), Cantonese subjects heard sentences presented at a fast rate and had to choose between two transcriptions of what they had heard; the transcriptions differed only in one character, representing a single difference of one syllable's vowel, tone, or vowel+tone. Accuracy was significantly greater for vowel differences than for tone differences, and vowel+tone differences were no more accurately distinguished than vowel differences alone. Repp and Lin (1990) asked Mandarin listeners to categorize nonword CV syllables according to consonant, vowel, or tone; categorization of tone was slower than categorization of vowel or consonant. Taft and Chen (1992) found that homophone judgments for written characters in both Mandarin and Cantonese were made less rapidly when the pronunciation of the two characters differed only in tone, as opposed to in vowel. Cutler and Chen (1997) similarly found that Cantonese listeners were significantly more likely erroneously to accept a nonword as a real word in auditory lexical decision when the nonword differed from a real word only in tone; and in a same-different judgment task, these listeners were slower and less accurate in their responses when two syllables differed only in tone, compared to when a segmental difference was present. In both tasks, an error was most probable when the correct tone of the real word and the erroneous tone on the nonword began similarly, in other words when the tone distinction was perceptually hard to make. Similar effects appear in the perception of Thai tones, in this case by non-native listeners: Burnham, Kirkwood, Luksaneeyanawin and Pansottee (1992) found that the order of difficulty of tone pairs presented in a same-different judgment task to English-speaking listeners was determined by the starting pitch of the tones.

Although tone contrasts are realized in F0, they are realized upon vowels, and therefore they are processed together with the vowel information. Yet vowels themselves can be identified very early; in a consonant-vowel sequence the transition from the consonant into the vowel is enough for listeners to achieve vowel identification (Strange, 1989). The evidence reviewed above suggests that tones can often not be identified so quickly - in speeded response tasks, subjects sometimes issue a response before the tonal information has effectively been processed. Thus although tone information is crucial for distinguishing between words in languages such as Cantonese, it may be the case that segmental information constrains initial lexical activation more strongly than tone information does.

Lexical Pitch Accent

Words in Japanese have patterns of pitch accent - high or low pitch levels associated with each mora of a polysyllabic word. Thus the word *Tokyo*, for example, has four

morae: *to-o-kyo-o,* of which the first has low pitch accent and the following three high, giving the word as a whole the pattern LHHH. Like lexical tone, pitch accent contrasts are realized via FO variation. There are quite a number of pairs of short Japanese words which differ only in accent pattern (such as *ame)* but very few such pairs of long words. Only a limited number of possible patterns exist. Japanese listeners find cross-spliced words with a correct segmental sequence but an impossible accent pattern (one which does not occur in the language) hard to process (Otake, Yoneyama, Cutler & van der Lugt, 1996).

Some recent experiments have suggested that Japanese listeners can make use of pitch accent information in early stages of word recognition, i.e. in the initial activation of word candidates. Cutler and Otake (1996) presented Japanese listeners with single syllables edited out of bisyllabic words differing in accent pattern, and asked them to judge, for each syllable, in which of two words it had originally been spoken. Thus the listeners might hear *ka* and be asked to choose between *baka* HL and *gaka* LH, or between *kage* HL and *kagi* LH; in other words, the listeners had to judge whether the syllable had H or L accent, since the syllable occurred in the same position in the two choice words, and the phonetic context adjacent to the *ka* was matched. The listeners performed this task with high accuracy, and their scores were significantly more accurate for initial (80% correct) than for final syllables (68%). This might suggest that pitch accent information is realized most clearly early in the word, where it would be of most use for listeners in on-line spoken-word recognition.

In a subsequent repetition priming study, Cutler and Otake (submitted) found that minimal pitch accent pairs such as *ame* HL and *ame* LH did not facilitate one another's recognition. Presentation of one member of the pair, in other words, apparently did not activate the other member, suggesting that a mismatch in pitch accent can rule out a candidate lexical item. In a gating study, the same authors presented listeners with successively larger fragments of words such as *nimotsu* HLL or *nimono* LHH - i.e. pairs of words with initial syllables (here, *nimo-)* having the same segmental structure but opposite pitch accent values. Listeners' incorrect guesses from about the end of the first vowel (ni-) overwhelmingly tended to be words with the same accent pattern as the actually spoken word. These results strongly suggest that Japanese pitch accent is exploited by listeners in word activation.

However, like lexical tone, pitch accent is realized via FO. and thus can only be reliably identified once a substantial part of the segment carrying it has been heard. In the gating study, the vowel in the first syllable constituted this necessary carrier segment. Walsh Dickey (1996) conducted a same-different judgment experiment in which Japanese listeners heard pairs of bisyllabic words or nonwords which were either the same, or differed either in pitch accent or in segmental structure. Just as Cutler and Chen (1997) observed for lexical tone. Walsh Dickey found that 'different' judgments were significantly slower for pairs varying in pitch accent than for

pairs which varied segmentally. Moreover, this was true irrespective of the position of the segmental difference; thus even a difference in a word-final vowel (at which time the pitch accent pattern of the whole bisyllable should be beyond doubt) led to significantly faster responses than the pitch accent difference.

Lexical Stress

Most experimental studies of word prosody have concerned lexical stress; and most of the research has been carried out in English. However, as this Section will outline, the role of lexical stress in word recognition may not be the same in English and in other stress languages.

In English, pairs of unrelated words differing only in stress pattern are rare; thus although stress oppositions between words of differing form class derived from the same stem *(import, contest)* are common, there are very few such pairs which are lexically clearly distinct (such *as forearm,* or *insight/incite).* Although stress could in principle provide minimal distinctions between words, in practice it rarely does so.

The rarity of minimal stress pairs is also true of German, Dutch and other languages with stress. However, the realization of stress in English differs somewhat even from other closely related languages. Unstressed syllables in English nearly always contain reduced vowels, and most full vowels bear either primary or secondary stress. This correlation is not nearly as strong in the other Germanic languages. The English word *cobra* for example has a reduced vowel - the vowel schwa - in the second syllable, where the equivalent German and Dutch words have the full vowel [a]; likewise, the English word *cigar* has schwa in the first syllable, while German *Zigarre* and Dutch *sigaar* have the full vowel [i]. Unstressed full vowels occur much more often in German and Dutch than they do in English.

The result of this crosslinguistic difference is that in English there are fewer pairs of words which can be distinguished suprasegmentally before they can be distinguished segmentally. Consider the words *alibi* and *alinea* (which exist both in German and in Dutch); both begin *ali-*, but in one the first syllable is stressed, in the other the second syllable. Such pairs practically do not exist in English; the initially-stressed word will almost certainly have schwa in the second syllable (this is true for instance of the word *alibi,* which does exist in English). Consequently, the earliest mismatching information which will become available to rule out a lexical candidate in English word recognition will virtually always be segmental information; the processing of suprasegmental information may make little useful contribution to constraining word activation. (In fact, statistical analyzes by Altmann and Carter [1989] established that the information value conveyed by phonetic segments in English is highest for vowels in stressed syllables.)

English listeners indeed find the distinction between full and reduced vowels more crucial than the distinction between stress levels; cross-splicing vowels with different stress patterns produces unacceptable results only if vowel quality is changed

(Fear, Cutler & Butterfield, 1995). In Fear et al.'s study, listeners heard tokens of words such as *audience*, which has primary stress on the initial vowel, and *audition*, which is one of the rare English words with an unstressed but unreduced initial vowel; when the initial vowels of these words had been exchanged, listeners rated the resulting tokens as insignificantly different from the original, unspliced, tokens. In a study of the recognition of words under noise-masking, Slowiaczek (1990) found that if vowel quality is not altered, mis-stressing has no significant effect on word identification. Changing vowel quality, on the other hand, does disrupt word recognition: thus Bond and Small (1983) found that mis-stressed words with vowel changes were not restored to correct stress in shadowing (indicating that subjects perceived the mis-stressed form and may not at all have accessed the intended word); and Bond (1981) found that the segmental distortion which could most adversely affect word recognition was changing full vowels to reduced and vice versa. A mis-stressing experiment by Cutler and Clifton (1984) similarly found a much stronger inhibitory effect of shifting stress in words with a reduced vowel *(wallet, saloon)* - since this necessarily involved a change in vowel quality - than in words with two full vowels *(nutmeg, canteen)*. Puns which involve a stress shift do not work (Lagerquist, 1980). Finally, a 'migration' experiment (in which phantom word recognitions are induced by combination of material presented separately to the two ears) by Mattys and Samuel (1997) demonstrated that mispronunciation in a stressed syllable inhibited construction of the phantom percept.

Knowing the stress pattern in advance does not facilitate word recognition in English: neither visual nor auditory lexical decision is speeded by prior specification of stress pattern (Cutler & Clifton, 1984). There are certain canonical correlations between stress pattern and word class in English (e.g., initial stress for bisyllabic nouns, final stress for bisyllabic verbs), and listeners know and can use this patterning in making 'off-line' decisions, i.e. responses that are not made under time pressure. Thus in studies by Kelly and colleagues (Cassidy & Kelly, 1991; Kelly, 1988, 1992; Kelly & Bock, 1988), subjects who were asked to use bisyllabic nonwords in a sentence as if they were words treated initially-stressed nonwords as nouns and finally-stressed nonwords as verbs; similarly, when asked to use a verb as a nonce-noun subject chose a verb with initial stress, while for a noun acting as a nonce-verb they chose a noun with final stress. However, this patterning again does not speed word recognition: whether or not a bisyllabic word conforms to the canonical pattern does not affect how rapidly its grammatical category is judged - *cigar* is perceived as a noun just as rapidly as *apple*, and *borrow* is perceived as a verb as rapidly as *arrive* (Cutler & Clifton, 1984).

In another off-line study, Connine, Clifton and Cutler (1987) asked listeners to categorise an ambiguous consonant (varying along a continuum between [d] and [t]) in either *DIgress-TIgress* (in which *tigress* is a real word) or *diGRESS-tiGRESS* (in which *digress* is a real word). Listeners' responses showed effects of stress-determined

lexical status, in that /t/ was reported more often for the *Dlgress-TIgress* continuum, but /d/ more often for the *diGRESS-tiGRESS* continuum. The listeners clearly could use the stress information in the signal, and in their stored representations of these words, to resolve the phonetic ambiguity. However, as with the correlation of stress pattern and word class, this off-line result can not shed light on the role of stress in on-line word activation.

If what matters for word recognition is primarily segmental identity, then the few minimal stress pairs in English, such *as forearm,* should be effectively homophones, just like all the many other English homophones *(match, count* etc.). Indeed, Cutler (1986) showed that this is so. In a cross-modal priming experiment (in which listeners hear a sentence and at some point during the sentence perform a visual lexical decision), Cutler found that both stress patterns, *FOREarm andforeARM,* facilitated recognition of words related to *each* of them (e.g., *elbow, prepare).* L. Slowiaczek (personal communication) similarly found priming for associates related to both phrase-stress and compound-stress readings of sequences such as *green house.* Thus English listeners apparently do not distinguish between two word forms distinguished only suprasegmentally in the process of achieving initial access to the lexicon; stress plays no role in on-line word activation.

As foreshadowed earlier, however, this state of affairs may hold for English only. The only other stress language for which a substantial body of experimental evidence exists is Dutch, but in Dutch, at least, the evidence now suggests a different picture. Van Heuven & Hagman (1988) analyzed a 70,000 word Dutch corpus to ascertain the contribution of stress to specifying word identity; they found that words could on average be identified after 80% of their phonemes (counting from word onset) had been considered; when stress information was included, however, a forward search was successful on average given only 66% of the phonemes. Off-line experiments in Dutch have demonstrated effects of stress on word identification. For instance, van Heuven (1988) and Jongenburger (1996) found that listeners could correctly select between two Dutch words with a segmentally identical but stress-differentiated initial syllable (e.g., *ORgel* and *orKEST,* or a minimal pair such as *SERvisch-serVIES)* when presented with only the first syllable. In a gating experiment, mis-stressing harms recognition, with mis-stressing of finally-stressed words *(PIloot* instead of *piLOOT)* more harmful than mis-stressing of initially-stressed words *(viRUS* instead of *Virus;* van Heuven, 1985; van Leyden & van Heuven, 1996; Koster & Cutler, 1997). Interestingly, another gating experiment by Jongenburger and van Heuven (1995a; see also Jongenburger, 1996), using minimal stress pairs (e.g., *VOORnaam-voorNAAM)* presented in a sentence context, found that listeners' word guesses only displayed correct stress judgments for the initial syllable of the target word once the whole of that initial syllable and part of the following vowel were available; this suggests that at least for minimal stress pairs, suprasegmental information may not exercise strong constraints on word activation.

Consistent with this, a cross-modal priming study in Dutch, planned as a direct replication of Cutler's (1986) experiment, failed to find any significant priming at all from initially-stressed members of stress pairs *(VOORnaam),* and inconsistent results for finally-stressed tokens *(voorNAAM;* Jongenburger & van Heuven, 1995b; Jongenburger, 1996).

Nonetheless, more recent results, using a larger population of words than is provided by the small set of minimal stress pairs, suggest that mis-stressing a Dutch word can prevent lexical activation. In word-spotting, embedded words are detected less rapidly when they occur within a string which itself could be continued to form a longer word; thus English *mess* is detected less rapidly in *doMES* than in *neMES,* presumably because *doMES* could be continued to form the word *domestic,* while *neMES* cannot be continued to form a longer real word (McQueen, Norris & Cutler, 1994). This finding replicates in Dutch: *zee* (sea) is harder to spot in *muzee* (which can be continued to form *museum)* than in *luzee.* However, if *muzee* is stressed not on the second syllable like *museum,* but on the initial syllable instead, i.e. listeners hear *MUzee* and *LUzee,* then there is no longer a significant difference between these in detection time for *zee* (Donselaar, Koster & Cutler, in preparation). This suggests that there was in this case no competition from *museum* because it simply was not activated by input lacking the correct stress pattern. Further, the fragment *aLI-* will prime *aLInea* but not *Alibi,* and the fragment *Ali-* will prime *Alibi* but not *aLInea* (Donselaar, Koster & Cutler, in preparation); this result was also observed with similar fragments of Spanish words (e.g., the first two syllables of *ARtico* or *arTIculo)* presented to Spanish listeners (Soto, Sebastian-Galles & Cutler, in preparation).

These last experiments in Dutch have not been attempted in English; can we in fact be sure that the same pattern of results would not after all show up in English with these new experimental methods? In fact, both experiments simply could not be replicated in English. The competition experiment *(zee* in *muzee)* requires words beginning with two strong syllables and containing a single embedded word; English does not contain sufficient numbers of such words. The fragment priming experiment *(all-* in *alibi* and *alinea)* likewise requires pairs of words beginning with two strong syllables; but equivalent words in English (such as *alibi)* contain a reduced vowel in one of the relevant syllables. The fact that such experiments are impossible in English is of course itself informative: it means that opportunities for listeners to use stress in the early stages of word recognition rarely occur in English, and words can virtually always be distinguished by segmental analysis without recourse to stress.

What then, can we conclude about the role of stress in word recognition? Indirectly, it of course plays a role due to the fact that stressed syllables are more acoustically reliable than unstressed syllables. Thus stressed syllables are more readily identified than unstressed syllables when cut out of their original context (Lieberman, 1963), and distortions of the speech signal are more likely to be detected in stressed

than in unstressed syllables (Cole et al., 1978; Cole & Jakimik, 1980; Browman, 1978; Bond & Games, 1980). In spontaneous speech, detection of word-initial target phonemes is also faster on lexically stressed than unstressed syllables (Mehta & Cutler, 1988); acoustic differences between stressed and unstressed syllables are relatively large in spontaneous speech, and such differences do not arise with laboratory-read materials. Stressed syllables are also recognized earlier than unstressed syllables in gated words, in spontaneously spoken but not in read materials (McAllister, 1991).

This does not imply that contrasts between stressed and unstressed syllables are salient to all listeners. Speakers of French, a language which does not distinguish words by stress, have great difficulty processing stress contrasts in nonsense materials, e.g., deciding whether a token *bopeLO* should be matched with an earlier token of *bopeLO* or *boPElo* (Dupoux, Pallier, Sebastian & Mehler, 1997). The same contrasts are easy for speakers of Spanish, a language which does distinguish words via stress. In fact, it should be noted that this entire Section has dealt with free-stress languages. There is as yet no direct evidence concerning the role of stress (e.g., the effects of mis-stressing) in fixed-stress languages, where contrasts between stressed and unstressed syllables exist but do not serve to distinguish one word from another. Indirect evidence is available from the word-spotting findings of Suomi, McQueen and Cutler (1997) in Finnish described above - the words excised from preceding monosyllabic contexts could be considered at least not to have their canonical stress, but no deleterious effects of this on word recognition were observed. Nevertheless there is room for new evidence from languages such as Finnish or Polish.

For free-stress languages, however, the evidence now suggests that stress may have a role in the initial activation of lexical entries in those languages where it contributes significant information to word identification; English is not one of these. Unfortunately, therefore, the language in which most psycholinguistic research (on any topic) is conducted turns out to be unrepresentative in the role its word prosody plays in word recognition.

2.4 Conclusion

The focus of the present chapter has been the process of recognizing spoken words and the ways in which prosodic structure directly influences that aspect of listening. There are of course much more general indirect influences which could have been considered. For instance, prosody plays a role in general intelligibility; just as sentences with acceptable syntax are understood more easily than sentences with abnormal syntax, sentences with plausible semantics are understood more easily than sentences with implausible semantics, and sentences with accurate phonetic realization are understood more easily than sentences with distorted phonetic structure, so are sentences with intact prosodic structure understood more easily than sentences

in which the prosodic structure has been in some way disrupted. This has been demonstrated in many languages (including German), and it is clear that word recognition would be one of the components of language processing affected by such manipulations. However, such general considerations fell outside the scope of this chapter.

The specific contributions of prosodic structure to word recognition, it has been argued, come on the one hand from sentence-level prosody - in which rhythm and grouping play a role in the discovery of word boundaries, and prominence can facilitate lexical processing - and on the other hand from the prosodic structure of words themselves, which, where it is suitably informative, is exploited to distinguish between candidate words. In all these research areas experimental evidence concerning the German language hardly exists, although evidence is indeed available from closely related languages. The review motivates the general conclusion that prosodic structure - like, one assumes, every other level of linguistic structure - is exploited by listeners in the process of word recognition to the extent that it can provide relevant and non-redundant information.

Acknowledgments

The overview given in this chapter overlaps to a considerable extent with the word-recognition Sections of the literature review compiled by Cutler, Dahan and van Donselaar (1997). I am very grateful to my co-authors Delphine Dahan and Wilma van Donselaar for joining me in that laborious but eventually rewarding project; the present chapter owes much to their advice and insights. I am also grateful to the publisher and editors of *Language and Speech* for permission to exploit further in this context the work done for the previous paper. Further thanks are due to Jens Bolte for assistance in tracking down studies of word recognition in German.

References

Altmann, G.T.M. & Carter, D.M. (1989). Lexical stress and lexical discriminability: Stressed syllables are more informative, but why? *Computer Speech and Language, 3,* 265-275.

Beach, CM. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30,* 644-663.

Birch, S. & Clifton, C.E. (1995). Focus, accent and argument structure: Effects on language comprehension. *Language and Speech, 38,* 365-391.

Birch, S.L. & Garnsey, S.M. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34,* 232-267.

Blutner, R. & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language, 27,* 359-367.

Bock, J.K. & Mazzella, J.R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition, II,* 64-76.

Bond, Z.S. (1981). Listening to elliptic speech: Pay attention to stressed vowels. *Journal of Phonetics, 9,* 89-96.

Bond, Z.S. & Games, S. (1980). Misperceptions of fluent speech. In R. Cole (ed.), *Perception and Production of Fluent Speech.* Hillsdale, NJ: Erlbaum.

Bond, Z.S. & Small, L.H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34,* 470-474.

Bradley. D.C., Sanchez-Casas, R.M. & Garcia-Albea, J.E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes, 8,* 197-234.

Browman, C.P. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics, 42.*

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C, Lacerda, F. & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 2514-2516). Philadelphia.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S. & Pansottee, S. (1992). Perception of Central Thai tones and segments by Thai and Australian adults. Pan-Asiatic Linguistics: *Proceedings of the Third International Symposium of Language and Linguistics* (pp. 546-560). Bangkok: Chulalongkorn University Press.

Buxton, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler & D.R. Ladd (eds.), *Prosody: Models and Measurements* . Heidelberg: Springer-Verlag.

Cassidy, K.W. & Kelly, M.H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*, 348-369.

Chen, H.-C. & Cutler, A. (1997). Auditory priming in spoken and printed word recognition. In H.-C. Chen (ed.), *The Cognitive Processing of Chinese and Related Asian Languages.* Hong Kong: Chinese University Press.

Ching, Y.C. (1985). Lipreading Cantonese with voice pitch. Paper presented to the 109th meeting, Acoustical Society of America, Austin *(Abstract Journal of the Acoustical Society of America, 77, Supplement 1*, 39-40).

Ching, Y.C. (1988). Voice pitch information for the deaf. *Proceedings of the First Asian-Pacific Regional Conference on Deafness* (pp. 340-343). Hong Kong.

Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*, 1570-1580.

Cole, R.A. & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America, 67,* 965-970.

Cole, R.A., Jakimik, J. & Cooper, W.E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64,* 44-56.

Collier, R. & Hart, J. (1975). The role of intonation in speech perception. In A. Cohen & S.G. Nooteboom (eds.), *Structure and Process in Speech Perception* . Heidelberg: Springer-Verlag.

Connine, CM., Clifton, C.E. & Cutler, A. (1987). Lexical stress effects on phonetic categorization. *Phonetica, 44,* 133-146.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics, 20*, 55-60.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201-220.

Cutler, A. (1987). Components of prosodic effects in speech recognition. *Proceedings of the Eleventh International Congress of Phonetic Sciences* (pp. 84-87). Tallinn, Estonia.

Cutler, A. (1991). Linguistic rhythm and speech segmentation. In J. Sundberg, L. Nord & R. Carlson (eds.), *Music, Language, Speech and Brain* . London: Macmillan.

Cutler, A. (1997). The syllable's role in the segmentation of stress languages. *Language and Cognitive Processes, 12,* 839-845.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,21*8-236.

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

Cutler, A. & Chen, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 106-109). Stockholm.

Cutler, A. & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59,* 165-179.

Cutler, A. & Clifton, C.E. (1984). The use of prosodic information in word recognition. In H. Bouma & D.G. Bouwhuis (eds.), *Attention and Performance X: Control of Language Processes* . Hillsdale, N.J.: Erlbaum.

Cutler, A., Dahan, D. & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141-201.

Cutler, A. & Darwin, C.J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29,* 217-224.

Cutler, A. & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7,* 49-59.

Cutler, A. & Foss, D.J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20,* 1-10.

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

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

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

Cutler, A., Norris, D.G. & Williams, J.N. (1987). A note on the role of phonological expectations in speech segmentation. *Journal of Memory and Language, 26,* 480-487.

Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33,* 824-844.

Cutler, A. & Otake, T. (1996). The processing of word prosody in Japanese. *Proceedings of the Sixth Australian International Conference on Speech Science and Technology* (pp. 599-604). Adelaide.

Cutler, A. & Otake, T. (submitted). Pitch accent in spoken-word recognition in Japanese.

Dahan, D. (1996). The role of rhythmic groups in the segmentation of continuous French speech. *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1185-1188). Philadelphia.

Dahan, D. & Bernard, J.M. (1996). Interspeaker variability in emphatic accent production in French. *Language and Speech, 39,* 341-374.

van Donselaar, W., Koster, M. & Cutler, A. (in preparation). *Voornaam* is not a homophone: Lexical prosody and lexical access in Dutch.

van Donselaar, W. & Lentz, J. (1994). The function of sentence accents and given/new information in speech processing: Different strategies for normal-hearing and hearing-impaired listeners? *Language and Speech, 37,* 375-391.

Dupoux, E. & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language, 29,* 316-335.

Dupoux, E., Pallier, C, Sebastian-Galles, N. & Mehler, J. (1997). Adestressing deafness in French? *Journal of Memory and Language, 36,* 406-421.

Eefting, W. (1991). The effect of'information value' and 'accentuation' on the duration of Dutch words, syllables and segments. *Journal of the Acoustical Society of America, 89,* 412- 424.

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.

Fox, R.A. & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics, 13,* 69-90.

van Heuven, V.J. (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America, 78,* 21.

van Heuven, V.J. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. *Proceedings of Speech '88, 7th FASE symposium* (pp. 811-818). Edinburgh.

van Heuven, V.J. & Hagman, P.J. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (eds.), *Linguistics in the Netherlands 1988.* Dordrecht: Foris.

Hohle, B. & Schriefers, H. (1995). Ambisyllabizitat imDeutschen: Psycholinguistische Evidenz. *Akten des 29. Linguistischen Kolloquiums.* Tubingen: Niemeyer.

Jongenburger, W. (1996). *The role of lexical stress during spoken-word processing.* Ph.D. thesis, Leiden.

Jongenburger, W. & van Heuven, V.J. (1995a). The role of linguistic stress in the time course of word recognition in stress-accent languages, *Proceedings of the Fourth European Conference on Speech Communication and Technology* (pp. 1695-1698). Madrid.

Jongenburger, W. & van Heuven, V.J. (1995b). The role of lexical stress in the recognition of spoken words: Prelexical or postlexical?, *Proceedings of the Thirteenth International Congress of Phonetic Sciences* (pp. 368-371). Stockholm.

Kakehi, K., Kato, K. & Kashino, M. (1996). Phoneme/syllable perception and the temporal structure of speech. In T. Otake & A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* . Berlin: Mouton de Gruyter.

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

Kelly, M.H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language, 27,* 343-358.

Kelly, M.H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349-364.

Kelly, M.H. & Bock, J.K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 389-403.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59,* 1208-1221.

Kolinsky, R., Morais, J. & Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language, 34,* 19-40.

Koopmans-van Beinum, F.J. & van Bergem, D.R. (1989). The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech. *Proceedings of the European Conference on Speech Communication and Technology* (pp. 113-116). Edinburgh.

Koster, M. & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of the Fifth European Conference on Speech Communication and Technology* (pp. 2167-2170). Rhodes.

Lagerquist, L.M. (1980). Linguistic evidence from paranomasia. *Papers from the Seventh Regional Meeting of the Chicago Linguistic Society,* 185-191.

Lee, Y.-S., Vakoch, D.A. & Wurm, L.H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25,* 527-542.

Lehiste, I., Olive, J.P. & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America, 60,* 1199-1202.

van Leyden, K. & van Heuven, V.J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In C. Cremers & M. den Dikken (eds.), *Linguistics in the Netherlands 1996* . Amsterdam: John Benjamins.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 172-187.

Martin, J.G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America, 65,* 1286-1297.

Mattys, S.L. & Samuel, A.G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language, 36,* 87-116.

McAllister, J. (1991). The processing of lexically stressed syllables in read and spontaneous speech. *Language and Speech, 34,* 1-26.

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

Mehta, G. & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech, 31,* 135-156.

Meltzer, R.H., Martin, J.G., Mills, C.B., Imhoff. D.L. & Zohar, D. (1976). Reaction time to temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 277-290.

Mens, L. & Povel, D. (1986). Evidence against a predictive role for rhythm in speech perception. *The Quarterly Journal of Experimental Psychology, 38A,* 177-192.

Nakatani, L.H. & Schaffer, J.A. (1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63,* 234-245.

Nooteboom, S.G., Brokx, J.P.L. & de Rooij, J.J. (1978). Contributions of prosody to speech perception. In W.J.M. Levelt & G.B. Flores d'Arcais (eds.), *Studies in the perception of language* . Chichester: John Wiley & Sons.

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

Otake, T,, Hatano, G. & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (eds.), *Phonological Structure and Language Processing: Cross-Linguistic Studies* . Berlin: Mouton de Gruyter.

Otake, T., Yoneyama, K., Cutler, A. & van der Lugt, A. (1996). The representation of Japanese moraic nasals. *Journal of the Acoustical Society of America, 100,* 3831-3842.

Pallier, C, Sebastian-Galles, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. Journal of *Memory and Language, 32,* 373-389.

Peretz, I., Lussier, I. & Beland, 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* . Berlin: Mouton de Gruyter.

de Pijper, J.R. & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America, 96,* 2037-2047.

Pitt, M. A. & Samuel, A.G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 564-573.

Quene, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics,* 20,331-350.

Quene, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America, 94,* 2027-2035.

Read, C. Kraak, A. & Boves, L. (1980). The interpretation of ambiguous who-questions in Dutch: The effect of intonation. In W. Zonneveld & F. Weerman (eds.), *Linguistics in the Netherlands 1977 -1979* . Dordrecht: Foris.

Repp, B.H. & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception. *Journal of Phonetics, 18,* 481-495.

Rietveld, A.C.M. (1980). Word boundaries in the French language. *Language and Speech, 23,* 289-296.

de Rooij, J.J. (1976). Perception of prosodic boundaries. *IPO Annual Progress Report, 11,* 20-24.

van Santen, J.P.H. & Olive, J.P. (1990). The analysis of contextual effects on segmental duration. *Computer Speech & Language, 4,* 359-390.

Schreuder, R. & Baayen, R. H. (1994). Prefix stripping re-revisited. *Journal of Memory and Language, 33,* 357-375.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America, 71,* 996-1007.

Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K. & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 375-380). Hillsdale, NJ: Erlbaum.

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

Shields, J.L., McHugh, A. & Martin, J.G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102,* 250-255.

Slowiazcek, L.M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech, 33,* 47-68.

Soto, S., Sebastian-Galles, N. & Cutler, A. (forthcoming). Stress and word recognition in Spanish.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85,* 2135-2153.

Streeter, L.A. (1978). Acoustic determinants of phrase boundary location. *Journal of the Acoustical Society of America, 64,* 1582-1592.

Suomi, K., McQueen, J.M. & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36,* 422-444.

Taft, M. & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O.J.L. Tzeng (eds.), *Language processing in Chinese* . Amsterdam: Elsevier.

Taft, M. & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language, 24,* 320-335.

Terken, J. & Nooteboom, S.G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*, 145-163.

Tsang, K.K. & Hoosain, R. (1979). Segmental phonemes and tonal phonemes in comprehension of Cantonese. *Psychologia, 22*, 222-224.

Tyler, L.K. & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26*, 638-657.

van Donselaar, W, Koster, M. & Cutler, A. (forthcoming) Lexical stress and lexical activation in Dutch.

Vroomen, J. & de Gelder, B. (1994). Speech segmentation in Dutch: No role for the syllable. *Proceedings of the Third International Conference on Spoken Language Processing*, Yokohama: Vol. 3, 1135-1138.

Vroomen, J., van Zon, M. & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition, 24*, 744-755.

Walsh Dickey, L. (1996) Limiting-domains in lexical access: Processing of lexical prosody. In M. Dickey & S. Tunstall (eds.), *University of Massachusetts Occasional Papers in Linguistics 19: Linguistics in the Laboratory* .

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P.J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91*, 1707-1717.

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 and Cognition, 19*, 260-271.