The Handbook of Speech Perception

Edited by

David B. Pisoni and
Robert E. Remez

Blackwell Publishing 2005

11 Lexical Stress

ANNE CUTLER

Stress is accentuation of syllables within words, or of words within sentences. This chapter deals with the first of these phenomena: lexical, or word stress. In lexical-stress languages, the syllables of any polysyllabic word are not created equal. Some syllables may serve as the locus of accentual prominence; others may not. Perceptually, this results in a distinction in salience between the syllables within a word. Thus *syllable* has initial stress: *SYLlable*. *Syllabic* is stressed on the second syllable: *sylLABic* (upper case denotes a stressed syllable).

Although the term *stress* is properly an abstraction, speech perception deals with physical realities, and so, as Section 11.1 describes, research on stress perception has largely been concerned with the acoustic characteristics of stressed versus unstressed syllables, and how listeners exploit the acoustic information to make decisions about where stress occurs. Differences across languages in the realization and function of stress have important perceptual consequences; this issue is considered in Section 11.2. Only in some languages is stress a potential contributor to spoken-word recognition; the empirical evidence on this question is surveyed in Section 11.3.

11.1 Acoustic Realization and Perceptual Apprehension of Stress

11.1.1 *The scope of this survey*

Like speech perception research in general, research on the perception of stress came of age when it became possible not only to measure the acoustic properties of speech signals, but to manipulate them. This was in the middle of the twentieth century.

The research described in this section concerns word stress in free-stress languages, although the manifestation of stress in free- versus fixed-stress languages has not always been kept apart in the literature (Section 11.2, below, will address this distinction). Also it has not always been the case that word stress has been kept apart from phrase and sentence stress, especially in earlier literature; however, prominence within the utterance more properly belongs in the domain of intonation research (see Vaissière, this volume).

In words, stressed syllables very often differ from unstressed syllables in nature – phonological systems assign stress to heavier syllables but not to lighter syllables, for instance. But stressed syllables also differ from unstressed syllables in their acoustic realization. Figure 11.1 shows spectrograms of the utterance *say the word pervert again*, spoken by a male speaker of American English. In the top spectrogram the verb reading (*perVERT*) is shown, in the bottom spectrogram the noun reading (*PERvert*); although the syllables have the same segmental structure in each reading, the acoustic realization is clearly different.

Speech perception research paid great attention to acoustic realization. The measurements and manipulations concerned the suprasegmental parameters of speech: that is, those dimensions within which any speech signal must be realized (Lehiste, 1970), but which in principle can vary independently while the segmental identity of a syllable remains constant. In practice this involved three acoustic dimensions: the durational patterning of the utterance, the fundamental frequency (F0) of the voice, and the signal amplitude. Perceptually these correspond to utterance timing, pitch, and loudness.

11.1.2 *Multiply determined stress judgments*

Fry (1955, 1958) conducted a systematic study of word stress in English minimal noun/verb pairs such as *OBject – obJECT*. Fry measured the duration and peak amplitude of each vowel, and found almost non-overlapping distributions for the stressed versus unstressed versions of comparable syllables, with the stressed syllables being longer and with higher amplitude than the unstressed. He then synthesized versions of the same words, varying these two parameters independently in 5 steps each, and varying F0 in 16 steps. Listeners' judgments of which syllable bore stress in the resulting synthetic words showed more effect of the durational manipulation than of the amplitude manipulation; change in F0 had an all-or-none effect, such that a syllable with a noticeably higher peak F0, or a clear F0 movement, was always judged to be stressed. The function of stressed syllables in stress languages is to serve as possible locations for accentual prominence within utterance intonation contours, and this result suggests that listeners are highly sensitive to where intonational prominence is realized; wherever this occurs, the location is judged to be a stressed syllable. Fry (1958) thus cautiously concluded that this F0 effect may be strongest for stress perception, with the effect of duration also being significant, but that of amplitude negligible. However, he also pointed out that vowel quality needed to be investigated, and he undertook this task in a subsequent study (Fry, 1965), in which shifts of the vowel formant ratios (e.g., for *object*, stepwise from [ɒ] to [ə] in the first syllable or from [ɛ] to [ɪ] in the second) were compared to durational and amplitude manipulations. The suprasegmental factors here proved more closely related to stress judgments than this type of vowel change.

If subjective impressions suggested that stressed syllables were louder than unstressed, this impression was quickly disconfirmed even by some of the earliest studies. Mol and Uhlenbeck (1956) reversed the amplitude relationship of stressed

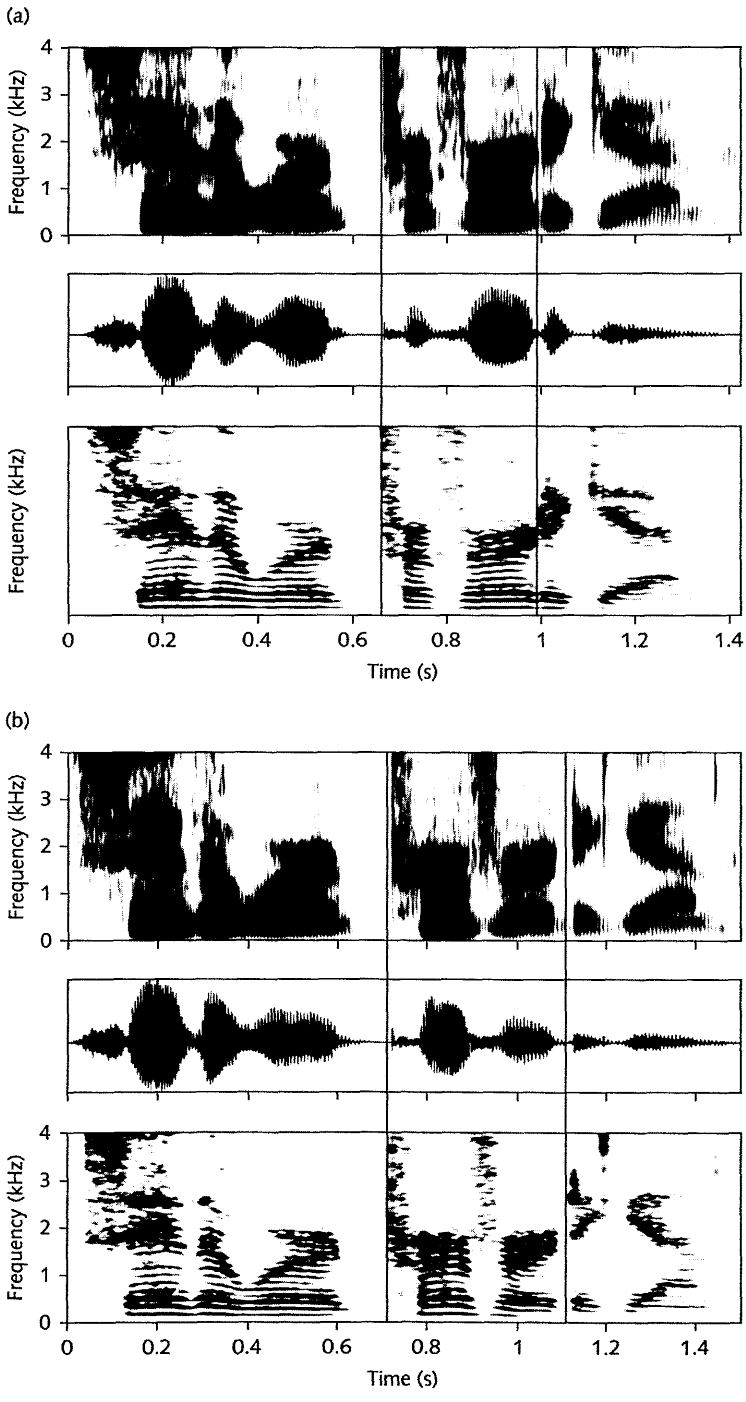

Figure 11.1 Sound spectrograms of the words perVERT (a) and PERvert (b), in the carrier sentence "Say the word . . . again," spoken by a male speaker of American English. Each figure consists of three display panels: above, a broad-band spectrogram; in the middle, a waveform display; and below, a narrow-band spectrogram. Vertical lines indicate onset and offset of pervert. The figure is modeled on a figure presented by Lehiste and Peterson (1959, p. 434).

versus unstressed syllables while leaving other parameters unchanged, and found that the reversal did not affect perceived stress; Bolinger (1958) added amplitude to pitch accents realized in synthesized sentences, and found that listeners' naturalness judgments if anything favored accents with less added amplitude. Ladefoged, Draper, and Whitteridge, introducing their pioneering 1958 electromyographic study of speech muscle activity, reported that "it is generally agreed" that stress has no single acoustic correlate (p. 9).

This was also the conclusion of studies by Lehiste and Peterson (1959), and Lieberman (1960). Lehiste and Peterson (on whose work Figure 11.1 is based) further drew attention to the non-independence of the acoustic parameters measured in stress studies and segmental factors – consonants and, especially, vowels can differ in intrinsic duration, amplitude, and pitch. Lieberman's study aimed to derive an algorithm for determining which of two syllables is stressed. Measurements of the three acoustic parameters in over 700 tokens of bisyllables (25 minimal pairs, produced in context by 16 speakers) showed that greater duration, higher average F0 and amplitude measures (higher peak, higher integral of amplitude across the syllable, and greater amplitude ratio of one syllable to the other in a word) were all strongly correlated with stress. Where one measure was not in the predicted direction, there was almost always a trade-off because other correlates of stress were present (in no fewer than 97% of cases). Lieberman concluded that no single acoustic cue to stress is necessarily important, but that all cues may be evaluated together.

11.1.3 *In search of a unitary underlying factor*

Should there be, no matter how complex the realization, some unitary underlying factor distinguishing stressed from unstressed syllables? The concept of articulatory effort figures in this role in many early studies (e.g. Fonagy, 1958, 1966; Ladefoged et al., 1958; Lehiste & Peterson, 1959; Van Katwijk, 1974). The latter two papers propose an explicitly perceptual account: "perception of linguistic stress is based upon judgements of the physiological effort involved in producing vowels" (Lehiste & Peterson, 1959, p. 428); "the perceptual effect [is] a pitch contour which could have been produced with an increment of subglottal pressure" (Van Katwijk, 1974, p. 66).

How could a notion like articulatory effort be tested? Lindblom (1963) considered that his measurements of Swedish vowel formant frequencies, which suggested invariant vowel targets attained to a greater or lesser extent as a function only of vowel duration, implied that vowel reduction should not be explained in terms of lesser articulatory effort: only timing patterns determined whether or not a vowel would be reduced. Van Katwijk's (1974) measurements of subglottal pressure found little evidence of stress-related pressure increase.

Harder still to assess was the proposal that listener judgments of stress depended on perceived effort. However a relevant contribution was made in the work of Isačenko and Schädlich (1966; followed up by Bleakley, 1973); in German utterances, stressed syllables could be signaled by any F0 obtrusion from the overall contour, so that a stressed syllable could be either higher or lower in pitch than its neighbors. Listeners rated both types of obtrusion as stress, which argued against commitment to greater perceived articulatory effort, and instead confirmed sensitivity to location of intonational prominence. Morton and Jassem (1965) found a similar result with English listeners judging synthesized nonsense bisyllables.

11.1.4 Reducing the complexity?

As the technology for manipulating speech signals improved, the complexity underlying stress perception was further confirmed. Nakatani and Aston (1978) used linear predictive coding techniques to manipulate orthogonally the natural duration, amplitude, and F0 attributes of bisyllables with initial or final stress, and found, as had others before them, that all exercised effects on stress judgments, although the effect of amplitude variation was weakest. In general, the effects were additive, except that durational variation lost its effect in sentence-final position, and F0 lost its effect when the word in question was deaccented in the intonation contour because it followed a contrastive accent (for comparable findings see Huss, 1975, 1978). Nakatani and Aston also reported speaker differences in the realization of stress contrasts (see also Howell, 1993). Other studies showed that differences between stressed and unstressed syllables in the three standard acoustic parameters were maintained across speech rates (Gay, 1978; McClean & Tiffany, 1973), while experiments in languages with fixed-position stress, or in languages with other phonological effects interacting with the stress system, further complicated the picture (see Section 11.2 below).

Attempts were made to develop perceptual measures that might facilitate greater comparability across studies. Gussenhoven and Blom (1978) proposed a "language-neutral test" based on paired judgments of perceived contrast between isolated vowels. Taylor and Wales (1987) proposed a contrast ratio:

$$(\text{stressed} - \text{unstressed})/(\text{stressed} + \text{unstressed})$$

which, they reported, for the three standard acoustic dimensions predicted judgments of perceived stress far more effectively than other ratios (such as the most commonly used subtraction ratio, i.e., the simple difference in any of the parameters between stressed and unstressed syllables, or the ratio stressed/unstressed, i.e., the division of one set of values by the other).

More recent research has tried to disentangle the complexity and multiple determination of stress perception by considering potential confounds. For instance, might stress judgments be complicated by the possibility that word stress placement can shift (so that English *thirteen* is stressed on the second syllable in *the number thirteen*, but on the first in *thirteen numbers*)? However, the undoubted shift in perceptual prominence in many such contrasting phrases does not result from a simple reversal of relative placement of a word's syllables on acoustic dimensions (see, e.g., Cooper & Eady, 1986; Shattuck-Hufnagel, Ostendorf, & Ross, 1994; Van Heuven, 1987). Likewise, measurements revealed little support for the notion of durational compensation within stress groups in English, such that segments of syllables might be longer in, say, a two-syllable foot (*save it*) than in a four-syllable foot (*savoring it*); segment durations are longer in syllables with primary stress than in syllables with secondary stress, and longer in turn in the latter than in unstressed syllables, such that syllable duration can be fully predicted given knowledge of the segments and stress pattern of the word (Crystal & House, 1988, 1990).

11.1.5 *Summary*

Research on stress perception continues both within and outside the speech perception literature. For example, it has long been of interest to speech engineers whether the use of stress-related information in the signal could improve the performance of automatic speech recognizers (see, e.g., Lea, 1977; Marshall & Nye, 1983; Waibel, 1988, for English; van Kuijk & Boves, 1999, for Dutch), and some successful implementations have been reported (e.g., Kiriakos & O'Shaughnessy, 1989; Sholicar & Fallside, 1988).

Over nearly 50 years, however, perceptual studies have elaborated but not fundamentally altered the early claims concerning the suprasegmental dimensions involved. Syllables are perceived to be stressed if they exhibit F0 excursion (Fry, 1958), whereby the timing of the F0 movement within the syllable can be crucial for determining stress perception (Thorsen, 1982, for Danish), and some types of F0 movement may require more excursion than others (Hermes & Rump, 1994, for Dutch F0 rise versus rise-fall). Greater syllable duration is likewise associated with perceived stress (Fry, 1955). These two factors are the most strongly related and the least controversial.

More controversial is the common finding that amplitude manipulations only weakly affect stress perception (despite psychoacoustic research showing that quite small changes in this dimension are indeed perceptible; Sorin, 1981). Turk and Sawusch (1996) found listeners' perception of duration and amplitude variation to be non-orthogonal; importantly, they observed effects of irrelevant durational variation on judgments of loudness to be greater than effects of amplitude variation on judgments of relative length. This, they argued, provided a rationale for why prominence judgments should be based on duration, or on duration and amplitude together, but not on the latter alone.

Beckman (1986) proposed that a measure of total amplitude (across a syllable) could capture effects on stress judgments, but as pointed out by Sluijter, van Heuven, and Pacilly (1997), such a measure is inevitably confounded with syllable duration. An indirect amplitude effect may however exist, in the factor which Sluijter and van Heuven (1996) termed spectral balance; stressed vowels have more energy in the higher frequency regions of the spectrum than unstressed vowels do. A linear discriminant analysis of their measurement data suggested that the most reliable correlate of the presence of stress was durational lengthening, with this spectral balance effect next in importance; overall amplitude had the usual weak impact on the analysis. A perceptual study of the spectral balance effect by Sluijter et al. (1997) showed that manipulations of this factor had a moderate effect when speech was presented to listeners over headphones, but a greatly increased effect when the speech was presented via loudspeakers! Sluijter et al. concluded that spectral balance directly reflects articulatory effort, in rehabilitation of the claim that perceived loudness is the most reliable cue to stress.

Campbell and Beckman (1997), however, failed to replicate the acoustic effect in English; they found spectral differences as a function of focal accent, but not as a function of lexical stress in the absence of accentual variation; this would therefore rule out analogous perceptual effects in English. On this issue, the last word may not yet be spoken.

11.2 Language-Specificity of Stress and Its Perceptual Consequences

11.2.1 Stress in phonological systems

The empirical evidence summarized in Section 11.1 was taken exclusively from studies of West Germanic languages. The evidence is similar across languages because the stress systems of these closely related languages are, although not identical, quite similar (see Figure 11.2). However, Germanic languages do not serve as a yardstick for languages across the world. First, stress characterizes the word-level phonology of only a subset of the world's languages. Second, there is considerable variation in word stress patterning even within stress languages, most obviously in the contrast between freely varying (as in the Germanic languages) and fixed-stress placement. Fixed-stress systems are considered in more detail below.

Little acoustic and perceptual evidence exists for non-Germanic free-stress languages. Measurements of segmentally matched stressed and unstressed syllables in Arabic by de Jong and Zawaydeh (1999) revealed duration and F0 correlates "remarkably like" the results reported for English (p. 20). In Spanish, stress is perceived if cued by F0 and duration or by F0 and amplitude, but not by any one cue alone (Llisterri et al., 2003); syllable weight and lexical analogy also affect stress perception (Face, 2000, 2003). Williams' (1985) experiments on synthesized Welsh minimal stress pairs found strong effects of duration on listeners' stress judgments, but inconsistent effects of F0. In Thai (a tone language), stress is signaled effectively by duration alone (Potisuk, Gandour, & Harper, 1996). It should be noted that some forms of English can also show other patterning than that described in Section 11.1. For instance, in Indian English (Bansal, 1966) and Welsh English (Williams, 1985) F0 movement can be decoupled from stress; in each case this can induce stress misperceptions by speakers of other varieties of English, who may for instance be led by F0 peaks or movements on unstressed syllables to judge those syllables as stressed.

In general, stress realization (and its perceptual reflection) will be dependent on other features of a language's phonological system. As Potisuk et al. (1996) point out, F0 does not vary as a function of stress in Thai because of its preemption by the tone system. Shen (1993) also observed that prominence in Mandarin is signaled by means other than F0. And just as tone preempts F0, so quantity distinctions in the segmental system render durational variation less useful for other purposes such as signaling stress. Berinstein (1979) found that speakers of Mayan languages with fixed final stress could learn to use duration variation as a position-independent cue to stress judgment, producing similar performance to

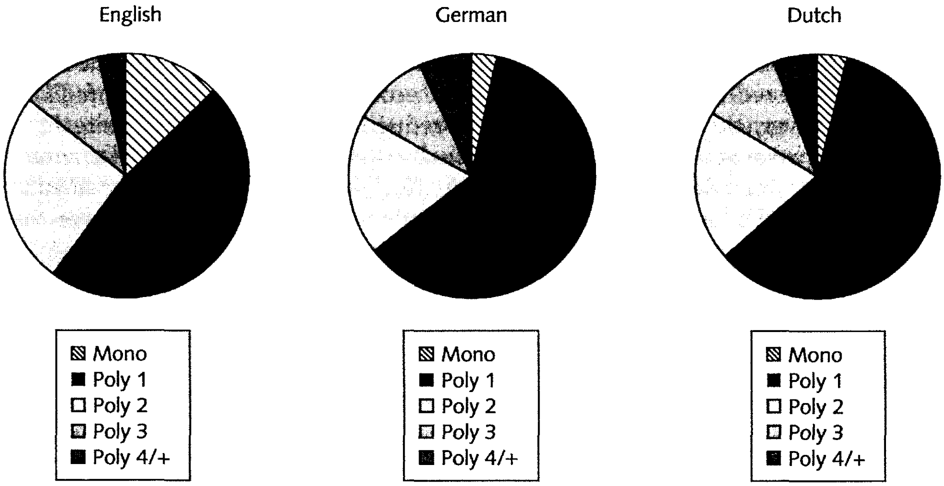

Figure 11.2 Distribution of lexical stress placement in three closely related languages: English, German, Dutch (from the CELEX lexical database; Baayen, Piepenbrock, & Van Rijn, 1993); monosyllabic words (Mono) versus polysyllabic words with stress on the first (Poly 1), second (Poly 2), third (Poly 3) or fourth or later syllable (Poly 4/+). The proportion of polysyllabic words is higher in the latter two languages; this is because the criteria for lemma inclusion admit many compounds in German and Dutch which are listed as separated lemmas in English. The tendency to word-initial stress in all three languages can be appreciated further when statistics for secondary stress are added. In English, secondary stress occurs on the initial syllable of about one-third of words with primary stress on the second syllable (7.7% in 25.2%), and most words with primary stress on the third (10.7% in 10.9%) or later syllable (2.3% in 3.6%). For German and Dutch, the values are similar: German 15.4% in 19%, 6.2% in 10.1%, and 3.5% in 6.8%; Dutch 12.7% in 20.1%, 8.1% in 10.5%, and 4% in 5.6%. Together, monosyllabic words and words with primary or secondary stress on the first syllable comprise 81% of the English lexicon and 89% of the German and Dutch lexicons.

speakers of English, but only if their language did not have vowel quantity distinctions – i.e., only if their language had not already preempted durational variation as a cue to something other than stress.

11.2.2 *Culminative vs. demarcative functions of stress*

Lexical stress variation has the word as its domain. In each word, only one syllable bears primary stress (although a few languages may not conform to this generalization; see Hyman, 1977, for further discussion). Thus stress is sometimes termed culminative within the word.

The single primary stress constraint does not rule out further distinctions among syllables which do not bear primary stress. Phonological accounts of the metrical structure of stress languages can encompass fine-grained distinctions of prominence within utterances (see Van der Hulst, 1999, for descriptions of the metrical structure of a range of European languages). Thus in English *PRESident*, primary stress falls on the initial syllable, whereas it falls on the third syllable in *presiDENtial*. But the remaining three syllables of *presiDENtial* are not all unstressed; the first syllable is said to bear secondary stress. The vowel in the initial syllable is not reduced, whereas those in the second and final syllables are likely to be; however, the possibility of vowel reduction is not a prerequisite for a contrast between secondary and lesser stress (consider Spanish *presidente*, which exhibits the same relative salience across its four syllables as English *presidential*, although Spanish has no vowel reduction). It is an empirical issue whether distinguishing between syllables with primary versus secondary stress (for example, in *presi-* from English *president* versus *presidential*) is perceptually necessary; this will be considered further in section 11.3 below.

The culminative function of stress may be contrasted with the so-called demarcative function, a term applying to stress which falls always at a particular position within the word. Again, one syllable within a polysyllabic word is the location of this fixed stress. The term demarcative refers to the potential for fixed-position stress to function as a marker of word boundaries (perhaps especially in languages with fixed word-initial stress, such as Finnish, Hungarian, Czech). Note, however, that fixed placement of stress of course implies that stress is not contrastive, i.e., cannot distinguish one word from another.

These properties are potentially important in perception. If stress can distinguish between words, listeners may use cues to stress in identifying spoken words; but if stress cannot help in this way, there is no reason for listeners to use it in word recognition. Similarly, if stress can signal word boundaries, listeners may use stress cues for segmenting continuous speech into words; but if stress were to have no systematic relation to position within the word, it could not be of use in segmentation.

11.2.3 *Free stress and segmentation*

Most research on the role of stress in the segmentation of speech has in fact been conducted in free-stress languages: notably in English and Dutch. Although both these languages are classified as having free stress, the place where stress will fall is not arbitrary, but is determined by considerations of syllable weight; morphological factors then conspire to place the designated stress-bearing syllable more often than not, as Figure 11.2 showed, in word-initial position. In typical conversational speech in English, the tendency to word-initial stress is even stronger than in the lexicon (Cutler & Carter, 1987). Listeners' behavior shows effects of this regularity; nonwords with initial stress can be repeated more rapidly, and attract higher word-likeness ratings, than nonwords with final stress (Vitevitch et al., 1997). The distributional asymmetry, combined with listener sensitivity to the pattern, opens the way for stress to be useful in segmentation in these languages, too.

Indeed, a substantial body of evidence from both English and Dutch indicates that listeners do treat stressed syllables as probable word onsets. Missegmentations of speech are more likely to involve stressed syllables being erroneously taken for word-initial and unstressed syllables being erroneously taken for word-internal than the reverse pattern, both in English (Cutler & Butterfield, 1992) and Dutch

(Vroomen, Van Zon, & De Gelder, 1996). Thus *a must to avoid* heard as *a muscular boy* is a natural error – the stressed last syllable is taken as a new word, while the unstressed two syllables preceding it are taken as internal to another word. Similarly, English listeners find word-spotting – detecting a real word in a spoken nonsense sequence – difficult if the word is spread over two strong syllables (e.g., *risk* in [rɪskib]) but easier if it is spread over a strong followed by a weak, unstressed syllable (e.g., *risk* in [rɪskəb]) (Cutler & Norris, 1988; see Quené & Koster, 1998, and Vroomen et al., 1996, for analogous evidence from Dutch). The difficulty in the former case is explained as resulting from division of the sequence at the onset of the second strong syllable on the strategy that any such syllable (with primary or secondary stress) is likely to be a new word; consequently, detection of *risk* in [rɪskib] requires that its component phonemes be reassembled, while no such delay affects detection of *risk* in [rɪskəb]. These findings suggest that listeners use distributional consistencies of stress placement to segment speech even in languages in which stress is not strictly demarcative. So far, however, similar studies of segmentation in fixed-stress languages are not available.

11.2.4 *Fixed stress discriminability*

In fact, there are reasons for caution with the superficially appealing notion of a demarcative role for stress cues in listening. One relevant consideration is that the acoustic realization of stress in fixed-stress languages is "weak" (Dogil, 1999; Rigault, 1970) in comparison to stress in free-stress languages. Early measurements of Hungarian (fixed initial stress) by Fonagy (1966) found the unstressed final syllables of bisyllabic words to be both longer and louder than the stressed initial syllables. Janota (1967) reported that F0 did not serve as a perceptual cue to stress in synthesized nonwords for Czech listeners (fixed initial stress); Rigault (1970) found in both Czech and French (fixed accent on final syllable of rhythmic groups) an absence of systematicity in stress realization. For Polish (fixed penultimate stress), Jassem (1962, cited in Morton & Jassem, 1965 and in Dogil, 1999) found no effective acoustic correlate of stress other than in F0. Dogil (1999) observed that Jassem's study was confounded with intonational variation, and conducted measurements of the same Polish words in different intonational frames; he found no consistent acoustic correlates of stress at all, and also no consistent reflections of putative rhythmic stresses in this language. Dogil proposed that in Polish, word stress has no other expression than as the abstract feature marking positions with which intonational movement may be associated. In contrast to lexical-stress languages, then, fixed-stress languages may not distinguish stressed from unstressed syllables at all in the absence of intonational realization of the abstract difference.

Another consideration is that although fixed stress is sometimes located at a word boundary, this is not always so (as in Polish, where stress is on the penultimate syllable). Given that the process of spoken-word recognition involves continuous exploitation of incoming acoustic information (see Frauenfelder & Floccia, 1998, for a review), fixed stress other than at a word edge will involve additional processing complexity. A third consideration is that fixed stress may affect all words irrespective of word class, or may be sensitive to grammatical factors.

This last consideration motivated investigations by Dupoux and Peperkamp (2002; Peperkamp & Dupoux, 2002), who aim ultimately to explain the acquisition of word prosody by infants. Dupoux and Peperkamp suggest that overall listener sensitivity to stress cues in adulthood will depend on the function of stress in their native language; in general, languages without lexically contrastive stress will not require that listeners develop sensitivity to stress information in speech. However, such non-contrastive languages differ in how accessible the rules for prosodic prominence are to an infant. In some languages (e.g., Finnish, French), prominence is unaffected by grammatical factors; infants should be able to learn early that stress plays no useful contrastive role and can be ignored. In other languages (e.g., Hungarian, Polish), prominence rules affect lexical words and grammatical (function) words differently; this pattern should be harder to learn, with the result that some sensitivity to stress contrasts may develop. Initial studies by Dupoux et al. (1997) and Dupoux, Peperkamp, and Sebastián-Gallés (2001) showed that speakers of French 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*. The same contrasts are easy for speakers of Spanish, which does distinguish words via stress. Dupoux and Peperkamp then tested Finnish speakers, whose performance indeed resembled that of French speakers, and speakers of Hungarian and Polish, whose performance fell between the French/Finnish and the Spanish levels.

11.2.5 *Fixed stress and segmentation*

Dupoux and Peperkamp's (2002) work suggests that fixed-stress languages cannot be treated as a unitary class. With regard to lexical segmentation, the available evidence on fixed-stress languages still falls far short of that on free-stress languages. In Indonesian (with phrase-final accent) listeners do not show evidence of using stress placement for segmentation (van Zanten & van Heuven, 1998). However, some studies on French and on Finnish (fixed initial stress) have provided evidence consistent with use of demarcation cues by listeners. In French, the accent-bearing right boundary of a rhythmic group is always also the right boundary of a word; Dahan (1996) found that listeners detect target syllables located at a rhythmic group boundary more rapidly than the same syllables elsewhere in an utterance. In Finnish, a word-spotting study by Suomi, McQueen, and Cutler (1997) showed that vowel harmony (which is a word-level phenomenon) can be used by listeners in segmentation; indirect evidence on stress processing can be deduced from their control experiment, in which excised embedded words from the word-spotting materials were recognized no less rapidly if taken from a preceding context (e.g., *palo* from *kupalo*) than from a following context (e.g., *palo* from *paloku*). Although the former type could be considered not to have been uttered with canonical stress, no deleterious effects of this on word recognition were observed. Vroomen, Tuomainen, and De Gelder (1998) replicated the Suomi et al. finding but showed that even stronger than vowel harmony was the effect on segmentation exercised by clearly marked word-initial stress. In a further experiment, Vroomen et al. showed that Finnish listeners were sensitive to vowel harmony and to stress (operationalized as higher F0) in learning the "words" of an artificial language, while Dutch listeners were sensitive only to the stress cue and French listeners were sensitive to neither cue.

These results are difficult to account for in the light of Dupoux and Peperkamp's (2002) finding that stress contrasts were overlooked by both French and Finnish listeners. However, it should be noted that the type of segmentation task used by Dahan (1996) does not directly tap word processing. Further, Vroomen et al. (1998) do not report how stress was realized in their word-spotting materials, but it is possible that as in their artificial language study, a clear F0 correlate was available. As Finnish has fixed initial stress, the initial syllable is the designated location for the realization of intonational prominence, but in the absence of such higher-level effects, Finnish stressed syllables are in fact not distinct in F0 from unstressed syllables (Suomi, Toivanen, & Ylitalo, 2003). The principal acoustic correlate of Finnish word stress is segmental lengthening within a word's initial two morae, even when the second mora is also the second syllable (Suomi & Ylitalo, 2004). The confound observed by Dogil (1999) in Jassem's (1962) work may thus also apply to Vroomen et al.'s study: listeners may have been able to use information relevant not to the lexical but to the intonational structure of the utterance (see Vaissière, this volume, for evidence on intonation perception).

11.2.6 *The contrastive potential of stress*

The putative contrastive function of stress must be considered in the light of the undeniable rarity of minimal word pairs which differ in stress alone (such as, for example, *trusty* and *trustee* in English). Free-stress languages make remarkably little use of the contrastive possibilities which stress in principle offers them. But there are other ways in which stress may be useful in word recognition. Statistical analyses by Altmann and Carter (1989) established that the amount of information conveyed by phonetic segments in English is highest for vowels in stressed syllables. Further, stressed syllables are acoustically reliable: they are more readily identified than unstressed syllables when excised from a context (Lieberman, 1963), and speech distortions are more likely to be detected in stressed than in unstressed syllables (Bond & Garnes, 1980; Browman, 1978; Cole & Jakimik, 1980; Cole, Jakimik, & Cooper, 1978). In gated presentation of spontaneously spoken – but not of read – sentences, stressed syllables are recognized earlier than unstressed syllables (McAllister, 1991). Also in spontaneous speech, word-initial target phonemes are detected more rapidly on lexically stressed than unstressed syllables (Mehta & Cutler, 1988). Note that acoustic differences between stressed and unstressed syllables are relatively large in spontaneous speech. With laboratory-read materials, however, such differences do not always arise; Mattys and Samuel (2000) found that phoneme detection was in general faster in words with initial stress, irrespective of whether the target phoneme occurred in the stressed syllable or elsewhere in the word.

Models of spoken-word recognition agree that continuous evaluation of speech input results in simultaneous activation of multiple candidate word forms which at any moment are supported by the input; eventual recognition proceeds on the basis of further input information but also via a process of competition between the activated words (Frauenfelder & Floccia, 1998). Within this presumably universal framework, the input information constraining activation will be necessarily language-specific. Like segmental contrasts, the relevant suprasegmental contrasts will differ across languages. If listeners do take account of stress, it can certainly help to reduce the number of word candidates. Van Heuven & Hagman's (1988) analyses of the Dutch vocabulary established 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 given only 66% of the phonemes. Wingfield, Goodglass, and Lindfield (1997) found that stress was relevant in determining the number of potential English word candidates from which listeners' recognition of gated words could be predicted. Section 11.3 considers empirical evidence for whether stress correlates in fact do constrain spoken word recognition.

11.3 Stress in the Recognition of Spoken Words

11.3.1 Lexical activation

The acoustic information in the signal which varies as a function of stress could play an early constraining role in lexical activation in the following way: as speech input activates word candidates, only those candidates which match the structure signaled by the input in stress as well as in segmental structure would become active. Words with non-matching stress or mismatching segments would not come into consideration.

Note that this means that stress information could play a substantial role in lexical activation even in languages where the number of word pairs distinguished by suprasegmental stress cues alone is vanishingly small. Word candidates may be activated by partial information as words are spoken; an utterance of *bottle* may temporarily cause activation of *bother, botch,* and *botany* among other words (Allopenna, Magnuson, & Tanenhaus, 1998; Zwitserlood, 1989). A word-initial syllable consisting of a given string of phonetic segments may be differently stressed in different words even though the words do not form a minimal stress pair. Although this is more likely in stress languages without vowel reduction, it also happens in English and similar languages. Thus *music* begins with stressed *mu-* while *museum* begins with unstressed *mu-*; *ad-* has primary stress in *admiral* but secondary stress in *admiration*. If stress cues distinguish these syllables for the purposes of lexical activation, then *mu-* is not the same syllable in *music* and *museum*, and the utterance of one will not activate, even temporarily, the other.

The same opposition can of course occur with non-initial portions of words – compare *-day* in *today* or *Tuesday*, or *-cide* in *decide* or *suicide* – and this difference may play a crucial role in listening situations in which initial portions of a word have for some reason not been heard. But the role of stress information in constraining lexical activation via distinctions in word-initial sequences is potentially even more significant.

Evidence from studies of the effect of segmental mismatch on lexical activation suggests that as incoming phonetic information matches one of two competitors but not the other, the losing competitor suffers inhibition (Vitevitch & Luce, 1998). Listeners exploit distinctive information rapidly to favor a matched competitor, which is thus enabled to compete more effectively and actually cause significant reduction in activation of its mismatched rival. The effects of a stress mismatch and a segmental mismatch were directly compared by Soto-Faraco, Sebastián-Gallés, and Cutler (2001). In their study, native speakers of Castilian Spanish heard spoken sentences (of a non-constraining type such as *He did not know how to write the word . . .*) ending with a word fragment which fully matched one of two potential words and differed from the other in just a single phoneme or in stress pattern. For instance, the fragment *prinCI-* (stressed on the second syllable) matches the first two syllables of the Spanish word *prinCIpio* (beginning) and differs only in stress from the first two syllables of the Spanish word *PRINcipe* (prince). Likewise, the fragment *sardi-* matches *sardina* (sardine) but mismatches *sardana* (a type of dance) in a single vowel, and the fragment *bofe-* fully matches *bofeton* (smack) but mismatches *boletin* (bulletin) in a single consonant. At the offset of the word fragment listeners saw a string of letters on a screen and were asked to decide whether this string was a real word. Their responses were significantly faster to a visually presented word after matching fragments (e.g., to SARDINA after *sardi-*, to PRINCIPIO after *prinCI-*, etc.) than after control fragments (e.g., *manti-*); responses after fragments which minimally mismatched and favored another word (e.g., to SARDINA after *sarda-*, to PRINCIPIO after *PRINci-*, etc.) were, crucially, significantly slower than responses after control fragments. The three types of mismatch information (vocalic, consonantal, stress) each produced the same pattern of inhibition in Soto-Faraco et al.'s experiment.

Donselaar, Koster, and Cutler (2005) replicated the stress comparison from Soto-Faraco et al.'s fragment priming study in Dutch, presenting fragments like *octo-* which matched one of either *OCtopus* or *okTOber* and mismatched the other only in stress placement. They too found that responses preceded by a matching prime were significantly facilitated, while responses preceded by a mismatching prime were slowed, in comparison to responses after the control prime. Again, listeners used the stress information to speed the victory of one of two competitors for lexical recognition. By comparison, a similar study by Cutler and Donselaar (2001) found that fragments like *MUzee* which mismatched *muSEum* in stress did not cause facilitation of the matched word, but also did not cause inhibition; no Dutch word begins *MUzee* so there is no competitor to inhibit the mismatched word. Words such as *museum* are thus activated only when their initial portions are appropriately stressed. This was shown directly by another of Cutler and Donselaar's (2001) experiments, using the word-spotting task, in which listeners monitor short nonsense strings for the presence of any embedded real word. McQueen, Norris, and Cutler (1994) had shown that this task could reveal competition effects; the English word *mess* was detected more rapidly in the nonsense context *neMES* (which activates no competitor) than in *doMES* (which activates *doMEStic*, competition from which slows the recognition of *mess*). Cutler and Donselaar replicated this result in Dutch: *zee* 'sea' was detected more rapidly in *luZEE* (activating no competitor) than in *muZEE* (activating *museum*). In *MUzee*, however, the detection of *zee* was not significantly slowed, suggesting that *museum* had not been activated.

A single-syllable fragment (e.g., the first syllable of *octopus* or *oktober*) produced significant facilitation if it matched, but did not produce inhibition if it mismatched, i.e., was taken from a word with the contrasting stress pattern (Donselaar et al., 2005); the same pattern appeared in a fragment priming study in German by Friedrich (2002). Friedrich, Kotz, and Gunter (2001) presented fragments of varying length, but did not use target pairs contrasting in onset stress such as *octopus-oktober*, so that (as in Donselaar et al.'s study with *museum*) no competition and hence no inhibition was involved; however, these authors found that facilitation did not significantly increase with increasing fragment size. The difference in inhibition due to the constraint in the competition process exercised by two syllables versus one is presumably a function of number of remaining potential competitors; after two syllables the competitor set will be smaller. In either case, cues to stress in the realization of word-initial portions are clearly used by listeners.

This series of studies on Spanish, Dutch, and German has brought the investigation of cues to stress into the currently accepted activation-competition framework of spoken-word recognition theory. Friedrich (2002) also measured evoked response potentials (ERPs), and found evidence for a difference in these measures between prime-target pairs which matched versus mismatched in F0 correlates of stress. It is to be expected that the coming years will see more studies of this issue using electrophysiological and brain imaging techniques. What is remarkable is that although, overall, most research on the perception of stress has been conducted on English, directly comparable experiments to those just described in Spanish, Dutch, and German have not been done in English. In fact there is good reason for this: directly analogous experiments are actually impossible. This is because of the strong tendency in English for any unstressed syllable adjacent to a stressed syllable to contain a reduced vowel. Thus there are effectively no such pairs in English as *octopus/oktober* in Dutch; the second syllable of English *octopus*, for instance, is reduced and hence has a different vowel than English *October*.

This does not however rule out partially comparable experiments. One possibility in English is to compare pairs in which the stress placement contrast does not involve primary stress on the first syllable versus primary stress on the second syllable, but another placement contrast. In fact Soto-Faraco et al.'s (2001) stress experiment included some pairs contrasting primary stress on second versus third syllables (e.g., *coMEdia* 'comedy' versus *comeDOR* 'dining room'), and so did Donselaar et al.'s (2005) Dutch fragment priming experiment (e.g., *dyNAmo* 'dynamo' versus *dynaMIET* 'dynamite'). These are again not possible to match exactly with English examples; but a first- vs. third-syllable contrast in primary stress can be achieved. There are in fact many English pairs in which the second syllable is reduced in both, but primary stress is either on the first or the third syllable – e.g., *admiral* versus *admiration*. In such pairs a fragment comprising only the first two syllables (e.g., *admi-*) would in one case have primary stress plus a weak syllable, in the other secondary stress plus a weak syllable.

Cooper, Cutler, and Wales (2002) carried out a fragment priming study using such English word pairs, and found clear evidence that English listeners too can make use of cues to stress in recognizing spoken words: *admi-* with primary stress on the first syllable activated ADMIRAL to a greater extent than ADMIRATION, while *admi-* with secondary stress on the first syllable activated ADMIRATION to a greater extent than ADMIRAL. Single-syllable fragments (e.g., *mus-* from *music* or *museum*) also produced facilitation when stress cues matched. Cooper et al. found, however, no evidence of inhibition from stress-mismatching primes, either with one- or two-syllable fragments; they argued that stress cues contribute less to resolution of inter-word competition in English than segmental information does, and in particular less than the wider range of stress contrasts contributes in Spanish and Dutch.

However, their results do show that English listeners can exploit suprasegmental stress cues in word recognition if they are given the opportunity. Cooper et al.'s conclusion thus modifies an earlier conclusion by Cutler (1986), reached on the basis of a priming study using not fragments but minimal stress pairs: for example *FORbear* versus *forBEAR*, or *trusty* versus *trustee*. Such pairs are rare but a few do exist in English. The task used was cross-modal priming of associated words: listeners were presented with sentences which were neutral until the occurrence of the critical pair, e.g.: *The person that she was hurrying to see was the trusty/ trustee . . .* , and made lexical decisions about words presented visually at offset of the critical word in the sentence. Whichever member of the stress pair had been heard, listeners' responses to associates of both members of the pair were facilitated in comparison to control words. Cutler argued that the undoubted suprasegmental differences between, for instance, *FORbear* and *forBEAR* were ineffective in constraining lexical activation, so that for English listeners *forbear* was effectively a homophone. L. Slowiaczek (personal communication) reached the same conclusion on the basis of the finding that phrase-stress and compound-stress realizations of sequences such as *green house* primed associates related to both. The findings of Cooper et al. (2002) suggest, however, that the proposed inutility of stress cues in the initial stages of lexical activation does not extend to all types of stress contrast in English. Note that a cross-modal priming study in Dutch, planned as a direct replication of Cutler's (1986) experiment, failed to find significant priming at all by initially-stressed members of minimal stress pairs (*VOORnaam* 'firstname'), and inconsistent results for finally-stressed tokens (*voorNAAM* 'respectable') (Jongenburger & van Heuven, 1995a; Jongenburger, 1996), despite the other clear evidence for the use of stress cues in activation in Dutch; studies of minimal pairs may thus not provide the best window on the exploitation of stress information.

11.3.2 Lexical selection

Other types of word recognition studies, using tasks which do not tap into the early activation and competition stages, also show that listeners exploit stress cues in distinguishing between spoken words. Connine, Clifton, and Cutler (1987), for example, asked listeners to categorize 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 *DIgress-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, this does not entail that knowing stress patterns in advance can facilitate access to the stored representations of words. Cutler and Clifton (1984) examined the effects of providing such information in a word recognition task, by comparing recognition of the same words in a blockedpresentation condition (all items presented were bisyllabic and initially stressed, for example) versus a mixed condition. They found that neither visual nor auditory lexical decision was speeded by prior specification of stress pattern.

Another way in which stress can play a role in word recognition is via canonical correlations between stress pattern and word class (e.g., initial stress for bisyllabic nouns in English, final stress for bisyllabic verbs; see Sereno, 1986, for the relevant statistics). Words which can be either nouns or verbs (such as *rescue* or *control*) show slight prosodic differences consistent with the canonical patterns when read in their two word class realizations (Sereno & Jongman, 1995). Listeners know and can use this; 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. Further, when asked to use verbs as nonce-nouns subjects chose verbs with initial stress, while for nouns acting as nonce-verbs they chose nouns with final stress. This is analogous to knowledge of other stress regularities which has been demonstrated in production studies (see Colombo, 1991, for a review).

However, canonical patterning again does not directly speed spoken-word recognition, so that, for instance, whether or not a bisyllabic word conforms to the dominant noun/verb 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). Arciuli and Cupples (2002) replicated this result. Davis and Kelly (1997) also found no significant difference in the same classification task for native English speakers, but interestingly nonnative speakers of English – whose responses were of course much slower than those of the native speakers – did show a response advantage in their study for words which conformed to the canonical stress pattern for nouns and verbs respectively.

Gating is a task in which words are presented in fragments of increasing size; the dependent variable is how large a fragment is needed for listeners to recognize the word. Jongenburger and van Heuven (1995b; see also Jongenburger, 1996), using Dutch minimal pairs such as *voornaam* in 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 result suggests again that minimal stress pairs may not exhibit the strongest possible effects of stress information on word activation, since it contrasts with another gating study (Van Heuven, 1988), in which listeners could correctly assign just the first syllable of a word, in sentence context, to one of two words in which it was respectively stressed versus unstressed (e.g., *si-* to *SIlo* versus *siGAAR*).

For English, Lindfield, Wingfield, and Goodglass (1999) conducted a gating study in which the presentation of word-initial fragments was contrasted with a condition in which the same fragments were presented along with additional information about how long the target word was, or how many syllables it had and what the stress pattern was. Recognition of the target occurred earlier in the latter condition. Arciuli and Cupples (2003) found however that adding low-pass filtered versions of the remainder of a gated word did not lead to earlier recognition compared with the gated fragment alone.

Related to gating studies are experiments in which listeners are presented with parts of words and asked to select (usually in a two-way forced choice) the source word. In Dutch, van Heuven (1988), Jongenburger (1996), and Cutler and Donselaar (2001) 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 *VOORnaam-voorNAAM*) when presented with only the first syllable; Cutler and Donselaar found that the second syllables of minimal stress pairs (e.g., *-naam*) could also be accurately judged.

The high proportion of correct responses in the Dutch studies (for example, 85% for first and 80% for second syllables in Cutler and Donselaar's experiment) was not equaled in a similar study in English by Mattys (2000), though here too listeners performed above chance with both two-syllable and one-syllable word-initial fragments (on average 62% and 54% correct, respectively). Similarly, Cooper et al. (2002) found that English listeners correctly assigned 59% of initial syllables to source words such as *music* versus *museum*. Remarkably, however, Dutch listeners outperformed the native speakers in the same experiment, scoring 72% correct assignments. As discussed in section 11.3.1, the contribution of stress information in resolving lexical competition may be greater in some other languages than in English, and this difference may allow some proficient non-native users of English to exploit English stress cues more effectively than native speakers do.

11.3.3 *Lexical mismatch*

Other evidence that English listeners do not make maximal use of stress information in speech comes from studies of the perception of mis-stressed words. Small, Simon, and Goldberg (1988) found that mis-stressing did not inhibit word recognition if it effectively created the target word's stress pair (e.g., *INsert* pronounced as *inSERT* or vice versa), though recognition was significantly inhibited if the mis-stressing created a nonword (e.g., *chemist* pronounced *cheMIST*, or *polite* pronounced *POlite*). Similarly, Bond and Small (1983) found that word recognition in shadowing was achieved despite mis-stressing as long as the mis-stressing did not result in an alteration of vowel quality; Slowiaczek (1990) found the same for word identification in noise. Cutler and Clifton (1984) found that shifting stress without altering vowel quality had a much smaller adverse effect on recognition than stress shifts which changed full vowels to reduced or vice versa.

In contrast, Dutch experiments on the perception of mis-stressed words (using gating: van Heuven, 1985; van Leyden & van Heuven, 1996; or a semantic judgment task: Cutler & Koster, 2000; Koster & Cutler, 1997) have shown that mis-stressing harms word recognition in that language, and at least in Koster and Cutler's (1997) study the effects of mis-stressing were of similar magnitude to the effects of segmental mispronunciation. Mis-stressing of finally-stressed Dutch words (*PIloot* instead of *piLOOT*) is more harmful than mis-stressing of initially-stressed words (*viRUS* instead of *VIrus*) both in gating (van Heuven, 1985; van Leyden & van Heuven, 1996) and in a semantic decision task (Koster & Cutler, 1997).

In German, the same result appeared when ERPs were recorded as listeners made decisions about correctly stressed versus mis-stressed words (Friedrich, 2002): *KAnal* instead of *kaNAL* produced a deviant electrophysiological response, while *kaNU* instead of *KAnu* did not.

The suggestion from these cross-linguistic comparisons is that mis-stressing is more harmful in other stress languages than in English. In English, deviant stress sometimes seems to have no effect at all. Thus, Slowiaczek (1991) presented listeners with a sentence context and a stress pattern and asked them to judge a target word for acceptability; she found that the stress pattern information was often ignored, in that listeners responded "yes" to words which were semantically acceptable in the context but did not have the target stress pattern.

A cross-splicing study by Fear, Cutler, and Butterfield (1995) suggested that listeners pay more attention to the distinction between full and reduced vowels than to stress distinctions among full syllables. Listeners in this study heard tokens of words such as *audience, auditorium, audition, addition*, in which the initial vowels had been exchanged between words; they rated cross-splicings among any of the first three of these as insignificantly different from the original, unspliced tokens. Lower ratings were received only by cross-splicings involving an exchange between the initial vowel of *addition* (which is reduced) and the initial vowel of any of the other three words. Especially the vowels in stressed syllables seem to be important to listeners. Bond (1981) compared the disruptive effects on word recognition of several types of segmental distortion; most disruptive was distortion of vowels in stressed syllables. The number of features involved in disruption of a stressed vowel is irrelevant; any replacement of such a vowel is harmful (Small & Squibb, 1989). Likewise, mispronunciations in stressed syllables inhibit phantom word recognitions resulting from the combination of dichotically presented input (Mattys & Samuel, 1997).

11.4 Conclusion

Unsurprisingly, given the distribution of psycholinguistic laboratories across the world, a majority of the research concerning the use of stress information in spoken-word recognition has been carried out in English. However, the role of lexical stress in word recognition may not be the same in English and in other free-stress languages. The evidence certainly suggests that stress cues play a role in the initial activation of lexical forms in those languages where it contributes significant information to word identification; Dutch, German, and Spanish all appear to be included in this category. English is less clearly a good member of the category; English listeners can use stress information in activation if given the opportunity, but the opportunity in fact arises less often in word discrimination in English than in the other languages studied. Furthermore, several experimental demonstrations of better use of English stress information by nonnative than by native speakers now exist. Thus the language in which most psycholinguistic research (on any topic) is conducted unfortunately turns out to be rather unrepresentative in the role its word prosody plays in word recognition.

Recommendations for future research are therefore obvious. More laboratory investigations of the role of stress in word recognition in other languages with lexical stress are needed. The application of new techniques to study lexical processing on line is also recommended. But it is not only in spoken-word recognition that the predominance of research on English has skewed the picture of lexical stress perception. The fact that most early studies were carried out in a language with free stress, i.e., in which stress can fall at different positions in different words, also determined expectations in later research. Stress is certainly the same across all stress languages in that it always refers to a distinction between those syllables which may express accentual prominence and those which may not; but its manifestations are different in free- versus fixed-stress languages. In English and similar languages, stressed syllables and unstressed syllables differ acoustically, and much research effort focused on the perceptual consequences of this. But in fixed-stress languages, as Section 11.2 described, such intrinsic acoustic differences between stressed and unstressed syllables are not necessarily to be expected. There are at least as many fixed- as free-stress languages in the world (Goedemans, 2003). In fixed-stress languages, too, far more perceptual research is needed. In all cases, effects due to intrinsic characteristics of stressed versus unstressed syllables must be distinguished from effects which arise from differences in the applicability of intonational prominence.

ACKNOWLEDGMENTS

The material in Section 11.3 was in large part covered in the review article of Cutler, Dahan, and Donselaar (1997), and I thank Delphine Dahan and Wilma van Donselaar once again for their collaboration on that project. Thanks also to Lalita Murty for valuable research for Section 11.1, Roel Smits, Ethan Cox, and Dennis Pasveer for assistance with constructing the figures, and Bob Ladd for helpful comments on the text.

REFERENCES

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–39.

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, 265–75.

Arciuli, J. & Cupples, L. (2002). The effects of lexical stress in visual and auditory word recognition. *Proceedings of the Ninth Australian International Conference on Speech Science & Technology*. Melbourne.

Arciuli, J. & Cupples, L. (2003). Stress typicality effects in native and non-native speakers of English. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2051–4). Barcelona.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Bansal, R. K. (1966). The intelligibility of Indian English. PhD thesis, London University.

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Dordrecht: Foris.

Berinstein, A. E. (1979). A crosslinguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics*, 47.

Bleakley, D. (1973). The effect of fundamental frequency variations on the perception of stress in German. *Phonetica*, 28, 42–59.

Bolinger, D. L. (1958). On intensity as a qualitative improvement of pitch accent. *Lingua*, 7, 175–82.

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

Bond, Z. S. & Garnes, S. (1980). Misperceptions of fluent speech. In R. Cole (ed.), *Perception and Production of Fluent Speech* (pp. 115–32). Hillsdale, NJ: Lawrence Erlbaum.

Bond, Z. S. & Small, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics*, 34, 470–4.

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

Campbell, N. & Beckman, M. E. (1997). Stress, prominence and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (eds.), *Intonation: Theory, Models and Applications (Proceedings of a European Speech Communication Association Workshop, September 18–20, 1997)* (pp. 67–70). Athens: ESCA and University of Athens Department of Informatics.

Cassidy, K. W. & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language*, 30, 348–69.

Cole, R. A. & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America*, 67, 965–70.

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.

Colombo, L. (1991). The role of lexical stress in word recognition and pronunciation. *Psychological Research*, 53, 71–9.

Connine, C. M., Clifton, C. E., & Cutler, A. (1987). Effects of lexical stress on phonetic categorization. *Phonetica*, 44, 133–46.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and nonnative listeners. *Language and Speech*, 45, 207–28.

Cooper, W. E. & Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language*, 25, 369–84.

Crystal, T. H. & House, A. S. (1988). Segmental duration in connected-speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, 4, 1574–85.

Crystal, T. H. & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88, 101–12.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201–20.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218–36.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–42.

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* (pp. 183–96). Hillsdale, NJ: Lawrence Erlbaum.

Cutler, A. & Donselaar, W. van (2001). *Voornaam* is not a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–95.

Cutler, A. & Koster, M. (2000). Stress and lexical activation in Dutch. *Proceedings of* *the Sixth International Conference on Spoken Language Processing*, vol. 1 (pp. 593–6). Beijing.

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, 113–21.

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

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. 1185–8). Philadelphia.

Davis, S. M. & Kelly, M. H. (1997). Knowledge of the English noun-verb stress difference by native and nonnative speakers. *Journal of Memory and Language*, 36, 445–60.

de Jong, K. & Zawaydeh, B. A. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, 27, 3–22.

Dogil, G. (1999). The phonetic manifestation of word stress in Lithuanian, Polish and German and Spanish. In H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 273–311). Berlin/New York: Mouton de Gruyter.

Donselaar, W. van, Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58A.

Dupoux, E. & Peperkamp, S. (2002). Fossil markers of language development: Phonological "deafnesses" in adult speech processing. In J. Durand & B. Laks (eds.), *Phonetics, Phonology, and Cognition* (pp. 168–90). Oxford: Oxford University Press.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French. *Journal of Memory and Language*, 36, 406–21.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness." *Journal of the Acoustical Society of America*, 110, 1606–18.

Face, T. L. (2000). The role of syllable weight in the perception of Spanish stress. In H. Campos, E. Herburger, A. Morales-Front, & T. J. Walsh (eds.), *Hispanic Linguistics at the Turn of the Millennium* (pp. 1–13). Somerville, MA: Cascadilla Press.

Face, T. L. (2003). Degrees of phonetic similarity and analogically-driven stress perception in Spanish. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1723–6). Barcelona.

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.

Fonagy, I. (1958). Elektrophysiologische Beiträge zur Akzentfrage [Electrophysiological contributions on accent]. *Phonetica*, 2, 12–58.

Fonagy, I. (1966). Electrophysiological and acoustic correlates of stress and stress perception. *Journal of Speech & Hearing Research*, 9, 231–44.

Frauenfelder, U. H. & Floccia, C. (1998). The recognition of spoken words. In A. D. Friederici (ed.), *Language Comprehension: A Biological Perspective* (pp. 1–40). Berlin: Springer Verlag.

Friedrich, C. K. (2002). Prosody and spoken word recognition: Behavioral and ERP correlates. PhD thesis. University of Leipzig.

Friedrich, C. K., Kotz, S. A., & Gunter, T. C. (2001). Event-related evidence of word fragment priming: New correlate for language processing? *Psychophysiology*, 38 (supplement 1), 42.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765–8.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126–52.

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. *Proceedings of the Fifth International Congress of Phonetic Sciences* (pp. 306–11). Basel/New York: Karger.

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech*, 21, 347–53.

Goedemans, R. (2003). A typology of stress systems. Unpublished manuscript, Leiden University.

Gussenhoven, C. & Blom, J. G. (1978). Perception of prominence by Dutch listeners. *Phonetica*, 35, 216–30.

Hermes, D. J. & Rump, H. H. (1994). Perception of prominence in speech intonation induced by rising and falling pitch movements. *Journal of the Acoustical Society of America*, 96, 83–92.

Howell, P. (1993). Cue trading in the production and perception of vowel stress. *Journal of the Acoustical Society of America*, 94, 2063–73.

Huss, V. (1975). Neutralisierung englischer Akzentunterschiede in der Nachkontur [Neutralization of post-nuclear accent contrasts in English]. *Phonetica*, 32, 278–91.

Huss, V. (1978). English word stress in the post-nuclear position. *Phonetica*, 35, 86–105.

Hyman, L. (1977). On the nature of linguistic stress. In L. Hyman (ed.), *Studies in Stress and Accent* (pp. 37–81). Southern California Occasional Papers in Linguistics, SCOPIL 4. Los Angeles: University of Southern California.

Isačenko, A. V. & Schädlich, H.-J. (1966). Untersuchungen über die deutsche Satzintonation [Studies on German sentence intonation]. In *Studia Grammatica VII* (pp. 7–67). Berlin: Akademie-Verlag.

Janota, P. (1967). An experiment concerning the perception of stress by Czech listeners. *Acta Univ. Carolinae, Philologica Phonetica Pragensia*, 6, 45–68.

Jassem, W. (1962). Akcent jezyka polskiego [Accent in the Polish language]. *Prace Jezykoznawcze*, 31 (Wroclaw).

Jongenburger, W. (1996). *The Role of Lexical Stress During Spoken-Word Processing*. PhD thesis, University of Leiden. The Hague: Holland Academic Graphics.

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. 1695–98). 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. 368–71). Stockholm.

Kelly, M. H. (1988). Phonological biases in grammatical category shifts. *Journal of Memory and Language*, 27, 343–58.

Kelly, M. H. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*, 99, 349–64.

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

Kiriakos, B. & O'Shaughnessy, D. (1989). Lexical stress detection in isolated English words. *Speech Communication*, 8, 113–24.

Koster, M. & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of EUROSPEECH 97* (pp. 2167–70). Rhodes.

Ladefoged, P., Draper, M. H., & Whitteridge, D. (1958). Syllables and stress. In *Miscellanea Phonetica III* (pp. 1–14). London: International Phonetic Association, University College.

Lea, W. A. (1977). Acoustic correlates of stress and juncture. In L. M. Hyman (ed.), *Studies in Stress and Accent* (pp. 83–119). Southern California Occasional Papers in Linguistics, SCOPIL 4. Los Angeles: University of Southern California.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT Press.

Lehiste, I. & Peterson, G. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, 31, 428–35.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, 33, 451–4.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6, 172–87.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773–81.

Lindfield, K. C., Wingfield, A., & Goodglass, H. (1999). The role of prosody in the mental lexicon. *Brain and Language*, 68, 312–17.

Llisterri, J., Machuca, M., de la Mota, C., Riera, M., & Rios, A. (2003). The perception of lexical stress in Spanish. *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2023–6). Barcelona.

Marshall, C. W. & Nye, P. W. (1983). Stress and vowel duration effects on syllable recognition. *Journal of the Acoustical Society of America*, 74, 433–43.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics*, 62, 253–65.

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.

Mattys, S. L. & Samuel, A. G. (2000). Implications of stress pattern differences in spoken word recognition. *Journal of Memory and Language*, 42, 571–96.

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

McClean, M. D. & Tiffany, W. R. (1973). The acoustic parameters of stress in relation to syllable position, speech loudness and rate. *Language and Speech*, 16, 283–90.

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, 621–38.

Mehta, G. & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech*, 31, 135–56.

Mol, H. G. & Uhlenbeck, G. M. (1956). The linguistic relevance of intensity in stress. *Lingua*, 5, 205–13.

Morton, J. & Jassem, W. (1965). Acoustic correlates of stress. *Language and Speech*, 8, 159–81.

Nakatani, L. H. & Aston, C. H. (1978). Perceiving stress patterns of words in sentences. *Journal of the Acoustical Society of America*, 63 (supplement 1), S55(A).

Peperkamp, S. & Dupoux, E. (2002). A typological study of stress "deafness." In C. Gussenhoven & N. L. Warner (eds.), *Papers in Laboratory Phonology VII* (pp. 203–40). Berlin: de Gruyter.

Potisuk, S., Gandour, J., & Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica*, 53, 200–20.

Quené, H. & Koster, M. L. (1998). Metrical segmentation in Dutch: Vowel quality or stress? *Language and Speech*, 41, 185–202.

Rigault, A. (1970). L'accent dans deux langues à accent fixe: Le français et le Tchèque [Stress in two fixed-stress languages: French and Czech]. *Studia Phonetica*, 3, 1–12.

Sereno, J. A. (1986). Stress pattern differentiation of form class in English. *Journal of the Acoustical Society of America*, 79, S36.

Sereno, J. A. & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, 38, 57–76.

Shattuck-Hufnagel, S., Ostendorf, M., & Ross, K. (1994). Stress shift and early pitch accent placement in lexical items in American English. *Journal of Phonetics*, 22, 357–88.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, 36, 415–33.

Sholicar, J. R. & Fallside, F. (1988). A prosodically and lexically constrained approach to continuous speech recognition. *Proceedings of the Second Australian International Conference on Speech Science and Technology* (pp. 106–11). Sydney.

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

Slowiaczek, L. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, 20, 465–81.

Sluijter, A. M. C. & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100, 2471–85.

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, 101, 503–13.

Small, L. H. & Squibb, K. D. (1989). Stressed vowel perception in word recognition. *Perceptual and Motor Skills*, 68, 179–85.

Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics*, 44, 272–80.

Sorin, C. (1981). Functions, roles and treatments of intensity in speech. *Journal of Phonetics*, 9, 359–74.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–32.

Suomi, K. & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, 32, 35–63.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36, 422–44.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, 31, 113–38.

Taylor, S. & Wales, R. (1987). Primitive mechanisms of accent perception. *Journal of Phonetics*, 15, 235–46.

Thorsen, N. (1982). On the variability in F0 patterning and the function of F0 timing in languages where pitch cues stress. *Phonetica*, 39, 302–16.

Turk, A. E. & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, 99, 3782–9.

van der Hulst, H. G. (1999). Word accent. In H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (pp. 3–116). Berlin/New York: Mouton de Gruyter.

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, S21.

van Heuven, V. J. (1987). Stress patterns in Dutch (compound) adjectives: Acoustic measurements and perception data. *Phonetica*, 44, 1–12.

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. 811–18). 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* (pp. 59–68). Dordrecht: Foris.

van Katwijk, A. (1974). *Accentuation in Dutch: An Experimental Linguistic Study.* Assen: Van Gorcum.

van Kuijk, D. & Boves, L. (1999). Acoustic characteristics of lexical stress in continuous telephone speech. *Speech Communication*, 27, 95–111.

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* (pp. 159–70). Amsterdam: John Benjamins.

van Zanten, E. A. & van Heuven, V. J. (1998). Word stress in Indonesian: Its communicative relevance. *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania* [Bijdragen tot de Taal-, Land- en Volkenkunde], 154, 129–47.

Vitevitch, M. S. & Luce, P. A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science*, 9, 325–9.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40, 47–62.

Vroomen, J., Tuomainen, J., & De Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, 38, 133–49.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, 24, 744–55.

Waibel, A. (1988). *Prosody and Speech Recognition*. London: Pitman.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, 13, 381–406.

Wingfield, A., Goodglass, H., & Lindfield, K. C. (1997). Word recognition from acoustic onsets and acoustic offsets: Effects of cohort size and syllabic stress. *Applied Psycholinguistics*, 18, 85–100.

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