Perception & Psychophysics
2000, 62 (2), 253-265

The perception of primary and secondary stress in English

SVEN L. MATTYS
State University of New York, Stony Brook, New York

Most models of word recognition concerned with prosody are based on a distinction between strong syllables (containing a full vowel) and weak syllables (containing a schwa). In these models, the possibility that listeners take advantage of finer grained prosodic distinctions, such as primary versus secondary stress, is usually rejected on the grounds that these two categories are not discriminable from each other without lexical information or normalization of the speaker's voice. In the present experiment, subjects were presented with word fragments that differed only by their degree of stress—namely, primary or secondary stress (e.g., /ˈprɑsɪ/ vs. /ˌprɑsɪ/). The task was to guess the origin of the fragment (e.g., "prosecutor" vs. "prosecution"). The results showed that guessing performance significantly exceeds the chance level, which indicates that making fine stress distinctions is possible without lexical information and with minimal speech normalization. This finding is discussed in the framework of prosody-based word recognition theories.

Scientists trying to model speech perception and word recognition are becoming increasingly interested in the role of prosody in decoding the speech signal. In particular, they are seeking to understand the incidence of *stress* in speech processing in domains as various as attentional processes in speech perception (e.g., Meltzer, Martin, Mills, Imhoff, & Zohar, 1976; Mens & Povel, 1986; Pitt & Samuel, 1990; Shields, McHugh, & Martin, 1974), word acquisition (e.g., Echols, Crowhurst, & Childers, 1997; Jusczyk, Cutler, & Redanz, 1993; Morgan, 1996), architecture of the speech system (e.g., Grosjean & Gee, 1987; Mattys & Samuel, 1997), and word segmentation (e.g., Cutler & Butterfield, 1992; Cutler & Norris, 1988).

The relationship between stress and segmentation constitutes an important research area, because word segmentation is one of the most difficult obstacles that speech engineers have to overcome to devise machines able to recognize spoken information. The speech segmentation problem originates from the fact that connected speech does not bear any obvious word boundary markers. Contrary to its written counterpart, speech is fairly continuous, with little acoustic information as to where words begin and where they end (Cole & Jakimik, 1980a, 1980b; Klatt, 1980; A. M. Liberman & Studdert-Kennedy, 1978). Such underspecification represents a challenge to speech machine designers, because lexical retrieval, in such conditions, is at best a heuristic matter (Marcus, 1984; Waibel, 1986).

The possibility that stress can help a listener parse the speech input into words was suggested early on in a study by Nooteboom, Brokx, and de Rooij (1978) and was instantiated by Nakatani and Schaffer (1978) with reiterant speech[1] (Carlson, Grandstrom, Lindblom, & Rapp, 1973). However, it was not until the past decade that stress-based approaches, backed up by both empirical and distributional data, became more fully incorporated into models of speech processing and segmentation. For example, a dominant notion in the literature on speech segmentation today is that strong syllables tend to be perceived as word onsets (Cutler & Butterfield, 1992; Cutler & Norris, 1988). This strategy is efficient, because most content words in everyday English are initial stressed (Cutler & Carter, 1987).

Although stress-based models appear to provide an effective way to solve the speech continuity problem, their actual efficiency depends mainly on the specific definition of stress on which they rely. A liberal definition of stress will allow a large number of syllables to serve as speech segmentors, whereas a more conservative definition will limit the number of such syllables. For those stress-based models that specify the type of stressed syllables that trigger segmentation, the definition is generally a liberal one. Cutler and Norris (1988) and Norris, McQueen, and Cutler (1995) include among strong syllables any syllable that contains a full vowel, whether this syllable is primary stressed (e.g., the first syllable in "decorative"), secondary stressed (e.g., the first syllable in "benediction"), or unstressed unreduced (e.g., the first syllable in "carbonic"). In this view, only reduced syllables (e.g., schwas) do not trigger segmentation. Such a dichotomous partition of stress is called *metrical*.

This work was supported by National Institute of Mental Health Grant R01 MH5166301. Special thanks are due Arty Samuel for his helpful suggestions on the issues discussed in this article and Robert Remez and two anonymous reviewers for their constructive comments. Thanks are also due Donna Kat for her help with computer programming and Debra Gimlin for proofreading the original manuscript. Correspondence concerning this article should be addressed to S. L. Mattys, House Ear Institute, Department of Communication Neuroscience, 2100 West Third Street, Los Angeles, CA 90057 (e-mail: smattys@hei.org).

—Accepted by previous editor, Myron L. Braunstein

With its liberal parsing criterion, metrical segmentation yields a high rate of correct word boundary detection. The proportion of English words starting with a full-vowel syllable—and, hence, the metrical segmentation hit rate—is estimated to be as high as 73%, and it is 90% when frequency of use is taken into consideration (Cutler & Carter, 1987).

The downside of a liberal stress-based segmentation strategy is that it also generates a substantial number of false detections of word boundaries. This drawback derives from the prediction that words containing several full vowels inevitably initiate segmentation in more than one location. Midword segmentation is bound to occur in at least some bisyllables (e.g., "migraine"), many trisyllables (e.g., "parasite"), and virtually all longer words (e.g., "controversy," "cosmopolitan"). Leaving aside the case of unstressed unreduced syllables, an inspection of over 65,000 words drawn from a computer-readable English phonetic dictionary (Moby Pronunciator) reveals that 21% of all bisyllables contain a primary and a secondary stressed syllables. This figure reaches 40% for trisyllables and an average of 70% for longer words. In all, 41% of all English words contain at least one secondary stressed syllable. This observation implies that each of these words theoretically undergoes at least two segmentation hypotheses, with one or more being incorrect. Obviously, these numbers become even greater when unstressed unreduced syllables are counted as segmentors as well, a feature typically assumed by stress-based models.

In theory, oversegmentation can be considerably reduced if, among the words activated by a strong syllable, priority is given to the longest ones (Cutler & Carter, 1987). Such restricted activation would prevent midword strong syllables from causing segmentation. However, an inconvenience of restricted activation is that most short words embedded in longer ones (e.g., "dip" in "diplomat") would fail to be activated on line and would, thus, require some sort of corrective mechanism. This aspect is of importance, because embeddedness turns out to be the rule more often than the exception. Luce (1986) reported that, when frequency of use was considered, over 60% of all words were embedded in longer words. Similarly, McQueen, Cutler, Briscoe, and Norris (1995) found that 84% of polysyllabic English words had at least one word embedded within them.

An alternative to metrical segmentation and, hence, to the high false detection rate (or FA rate) it produces consists of adopting a more conservative approach to stress as a word boundary marker. In this case, functional distinctions are made between, for instance, primary stressed, secondary stressed, and unstressed unreduced syllables. From here on, two scenarios are possible. One possibility is that segmentation is attempted on, say, primary stressed syllables only. The other syllables, whether they bear a full syllable or not, would not prompt segmentation. The other possibility is that segmentation is initiated *more or less*, as a function of stress degree. That is, the higher a syllable's degree of stress, the higher the likelihood that the input will be segmented—or the heavier the weight of the activated words in the lexical access process.

In these two implementations, the number of false detections of word boundaries is considerably reduced, either because there are fewer syllables capable of segmenting the input or because syllables are less capable of doing it. For example, a system that segments speech on primary stressed syllables only would produce a hit rate of 88% (when the words are frequency weighted) and an FA rate of 12% (in this case, the FA rate also corresponds to the percentage of missed word boundaries). The hit and FA rates are derived for different word lengths in Figure 1, which displays the distribution of content words in English (panel A, without frequency weight; panel B, frequency weighted). This figure shows that the candidates for proper segmentation—that is, words starting with a primary stressed syllable—are in the majority, and overwhelmingly so when they are weighted by their frequency of occurrence in the language. For heavily represented word length categories, the hit rate of a primary-stress segmentation strategy is almost as high as that of a metrical strategy. The FA rate, on the other hand, is lower with a primary-stress strategy than with a metrical strategy in all word length categories (except for monosyllables that do not generate FAs). Indeed, FAs in primary-stress models are limited to those words starting with a secondary stressed or unstressed syllable. In metrical models, FAs are found in every word containing a secondary stressed syllable.

Alternatively, a model that links segmentation strength to degree of stress would generate a fairly good approximation of the descending probabilities that a syllable begins a word: 54% for primary stressed syllables, 18% for secondary stressed syllables, 21% for syllables containing an unstressed unreduced vowel, and 7% for syllables containing a reduced vowel (a schwa). Note that the last category becomes somewhat inflated when words starting with a syllable containing a reduced vowel other than a schwa (e.g., "invest," "external") are included in the count. Conversely, the percentage of unstressed unreduced syllables is expected to be noticeably lower in conversational speech, where vowels tend to be reduced to schwas.

One aspect shared by the latter two implementations is that syllable classification goes beyond a simple contrast between full and reduced vowels. Here, differences in stress levels are not necessarily accompanied by differences in vowel quality. Rather, it is suprasegmental variables, such as frequency, duration, and amplitude, that determine stress perception. In line with this approach, several studies have recently reported data suggesting that listeners might rely on such subtle stress differences to segment words from fluent speech (e.g., Vroomen & de Gelder, 1997; Vroomen, Tuomainen, & de Gelder, 1998). Vroomen and de Gelder found that Dutch listeners use a stress-based segmentation (SBS) strategy—whereby word boundaries are better signaled by primary than by secondary stressed syllables—rather than a metrical strat-

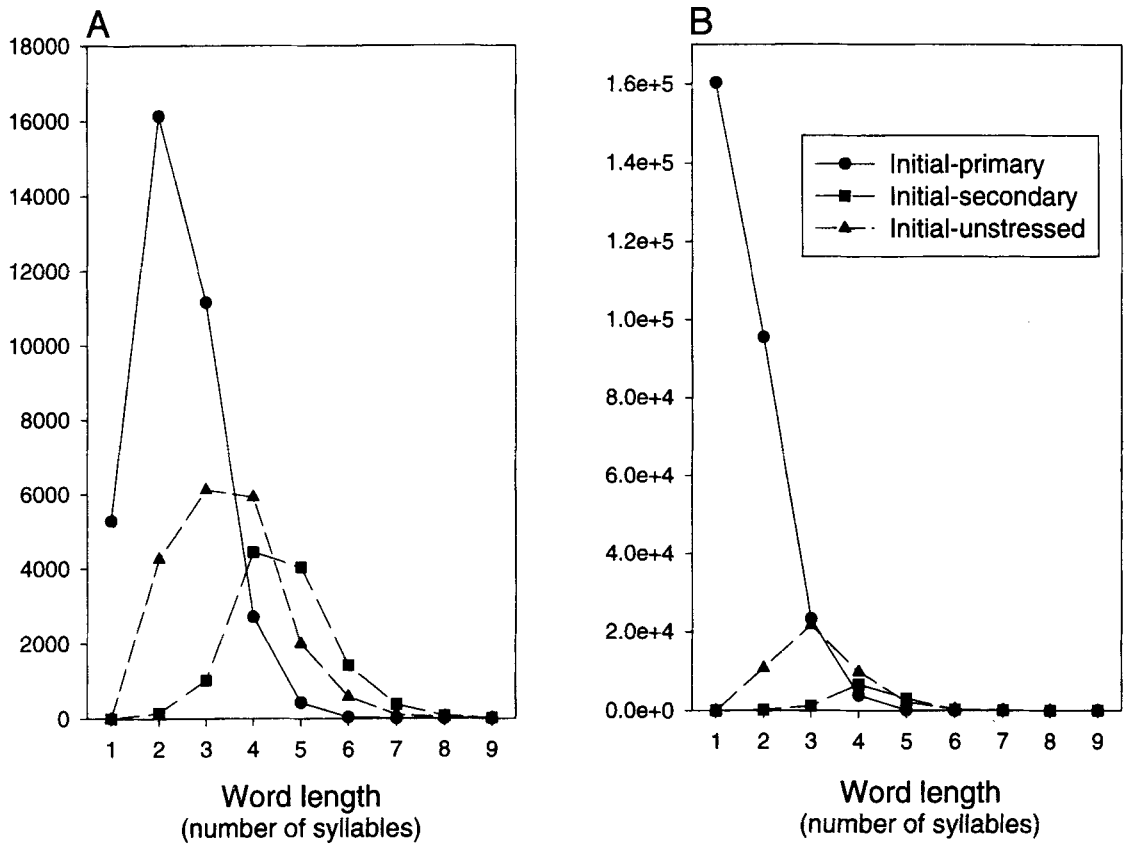

Figure 1. Distribution of initial-primary-stress, initial-secondary-stress, and initial-unstressed content words in English for nine word lengths in (A) absolute values and (B) frequency-weighted values. The distributional figures are obtained from Moby Pronunciator, and the frequencies from Kučera and Francis (1967).

egy. In one of the experiments, Dutch subjects had to detect a bisyllabic trochaic word (e.g., "CRAter") embedded in a trisyllabic string. The results showed that, even after the acoustic differences were factored out of the response times, the target word was detected faster when its initial syllable was realized as a primary stressed syllable in the string (e.g., /"pɔ'krɑtər/) than when it was secondary stressed (e.g., /'pɔ"krɑtər/). This result suggests that segmentation is guided more by the degree of a syllable's stress than by its vowel quality. A metrical segmentation strategy, which attributes the same segmentation power to any syllable bearing a full vowel, would have predicted that "CRAter" would be segmented from the two strings equally quickly.

Several studies of speech perception in infants also provide support for the notion that stress distinctions that do not affect vowel quality can influence perceptual preferences and, presumably, constrain segmentation strategies (e.g., Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996). For instance, 9-month-olds were found to prefer long/short bisyllables over short/long ones (Morgan, 1996) and stressed/unstressed over unstressed/stressed ones (Mattys et al., 1999), despite the fact that, in both cases, the first *and* the second syllables of the stimuli contained a full vowel. In other words, infants exhibited perceptual responses that were more compatible with a fine-grained stress segmentation strategy (e.g., primary vs. secondary stress) than with a metrical segmentation strategy.

The goal of the present paper is not to establish whether metrical segmentation or a more restricted SBS approach is correct but only to evaluate a necessary condition under which a restricted stress-based approach could be envisioned—namely, high stress-discriminatory perceptual capacities. If the speech processor does indeed initiate segmentation solely on primary stressed syllables or if the probability that a syllable is used to segment speech is a function of its degree of stress, we need to demonstrate that listeners have the perceptual capacity to discriminate between levels of stress that do not entail a change in vowel quality. Without evidence that they do, it is pointless to elaborate theories of speech segmentation in which fine distinctions between stress levels are made.

When it comes to assessing whether listeners can distinguish between different degrees of stress with or without minimal contextual information, the literature offers an unclear picture. On the one hand, researchers have failed to show that people can perceive anything but vowel

quality (i.e., full vs. reduced vowels), which argues against the fine prosodic segmentation hypothesis. For instance, Lieberman (1965) used electronically altered speech that, on output, sounded like strings of repeated /a/s mimicking the pitch and amplitude of the original signal (a sentence). He then analyzed the stress transcriptions produced by a linguist on both the original and the altered speech. The linguist's transcript of the real speech showed that he could discriminate between four levels of stress. However, in the transcript of the altered passages, he could distinguish only between stressed and unstressed syllables. Secondary and tertiary stresses were almost never correctly identified (7% and 0% intertranscript reliability). Despite the indication that the linguist had more difficulty being consistent in labeling fine stress categories, the interpretation can hardly be generalized to a larger population. In addition, these stress judgment data were obtained incidentally, since the subjects (two linguists, originally) were instructed to transcribe only the pitch contour of the passages. Only 1 subject reported stress estimates. Furthermore, stress per se was not manipulated in the design.

A more informative piece of research on stress processing can be found in Fear, Cutler, and Butterfield's (1995) data on word acceptability judgments. In their experiments, the authors had subjects rate the naturalness of words produced in sentences. The words, which all started with a vowel, were of four sorts, according to their initial syllable. The initial syllable could be (1) primary stressed (e.g., "*au*tumn"), (2) secondary stressed (e.g., "*au*tomation"), (3) unstressed unreduced (e.g., "*au*tomata"), and (4) reduced (e.g., "*a*tomic"). In the experiment, the test vowels were exchanged through cross-splicing, thus generating a four-vowel by four-condition design. The results showed that word acceptability rating was sensitive to cross-splicing only when the exchange was performed between any of the three full vowels and a reduced vowel. Cross-splicing within the full-vowel category did not significantly affect the acceptability ratings. The authors interpreted their results as an indication that listeners do distinguish between full and reduced vowels but presumably do not (or do less so) between intermediate categories.

It is important to point out, however, that these acceptability results cannot be directly equated with perceptual sensitivity. That is, an absence of rating difference between two categories (e.g., primary and secondary stress) does not necessarily mean that subjects are unable to perceive the difference between the two categories. In keeping with Cutler (1986), dichotomous stress categorization may suffice to identify most words unambiguously—hence, the absence of acceptability differences—whereas higher perceptual sensitivity to stress differences could be independently exploited to locate word boundaries.

Perceptual sensitivity to stress differences was investigated in a pioneering study by Fry (1958). Fry (1958) demonstrated that it was possible to make subjects judge a sequence such as /dɑɪdʒest/, which has two full syllables, to be either the noun "DIgest" or the verb "diGEST" by modulating the relative frequency, duration, or amplitude of the two syllables. Although this result in itself does not directly demonstrate that listeners are able to discriminate fine stress differences in natural speech, it shows that stress perception goes beyond a simple evaluation of vowel quality (in Fry's, 1958, experiment, the vowels' formant structure was always preserved).

Generally, listeners are fairly good at detecting small differences in fundamental frequency (F_0), duration, and amplitude, which are three acoustic dimensions traditionally associated with stress perception (e.g., Fry, 1955, 1958; Lehiste, 1970; Morton & Jassem, 1965). Research has indicated that subjects can discriminate bisyllables differing slightly in their syllables' respective F_0 (Lieberman, 1960; Morton & Jassem, 1965), as well as in pitch movements (Hermes & Rump, 1994; Hermes & Van Gestel, 1991; Pierrehumbert, 1979). Since the changes under investigation do not typically involve vowel reduction, the F_0 range roughly falls within—or overlaps with—that entailed in primary/secondary stress distinctions. This suggests that listeners may have the perceptual capacity to exploit such sensitivity to judge whether a syllable bears primary or secondary stress. Small changes in vowel duration and/or amplitude are also easily distinguishable (Brandt, Ruder, & Shipp, 1969; Sluijter & van Heuven, 1996; Sluijter, van Heuven, & Pacilly, 1997). For example, Sluijter et al. presented listeners with the bisyllabic reiterant sequence /nana/ and had them report, in a two-alternative forced-choice task, the syllable on which prominent stress was placed. The results showed that the percentage of *word-initial stress* responses decreased in a fairly monotonic way along the seven steps between long–short stimuli (250 msec ~ 185 msec) and short–long stimuli (130 msec ~ 275 msec). A similar response curve was obtained when the spectral balance (an index of amplitude based on the intensity ratio between high- and low-frequency bands) was manipulated from loud–soft (+3 dB ~ baseline) to soft–loud (baseline ~ +3 dB). These results indicate that listeners are sensitive to variables that usually participate in the primary/secondary stress distinction. Specifically, stress perception was found to evolve smoothly over small increments or decrements in duration (20 msec) and amplitude (1 dB), which correspond to the ranges of acoustic differences between primary and secondary stressed syllables in the present experiment (but see Fear et al., 1995, for partially discrepant acoustic data).

The incidence of the primary/secondary distinction in word recognition was apparent in a recent study by Mattys and Samuel (in press). In one of their experiments, the authors measured syllable-monitoring times in four-syllable-long words that either bore primary stress in the first syllable and secondary stress in the third syllable (e.g., "generator") or bore secondary stress in the first syllable and primary stress in the third syllable (e.g., "panorama"). The critical syllable to monitor was located between the two stressed syllables of the test words and was always identical within a pair (e.g., /nə/ in the example above). The data revealed that reducing the

availability of the third and fourth syllables of the test words by delaying them (a 200-msec pause was inserted after the target syllable) and covering them with noise was more detrimental to initial secondary stressed words (e.g., "panorama") than to initial primary stressed words (e.g., "generator"). That is, compared with the monitoring RTs in the intact words, degradation of the third and fourth syllables of "panorama"- and "generator"-type words slowed down the detection of the second syllable /nə/ to a greater extent in the former than in the latter case. For the authors, the disadvantage for initial secondary stressed words resulted from the reduction of the accessibility of the primary stressed syllable induced by the manipulation. The results were thus taken to be an indication that primary stressed syllables are more important than secondary stressed syllables in lexical access. Had the primary and the secondary stressed syllables been treated the same way, delay/noise degradation should have affected both word types equally.

What this finding suggests is that such subtle acoustic distinctions as those between primary and secondary stress might be picked up by the speech processor to guide word recognition. However, to validate this interpretation, it is important to show that listeners have the capacity to discriminate between primary and secondary stressed syllables when segmental information is kept constant.

The goal of this study is to explore the perceptual sensitivity to primary versus secondary stress. In order to control for segmental information without having to resort to signal distortion (e.g., low-pass filtering), the design included intact word fragments that differed only in the degree of stress in the first syllable (e.g., /ˈprɑsɪ/ or /ˌprɑsɪ/). The two fragments originated from real words recorded from naive speakers (e.g., "prosecutor" and "prosecution," respectively, for the example above). The listener's task was to guess which one of the two words the fragment originated from. Moreover, in an attempt to establish how the size of the word fragment relates to the performance, the subjects in another condition were presented with only the first syllable of the words (e.g., /ˈprɑ/ or /ˌprɑ/), and the accuracy of their guess was measured.

METHOD

Subjects

Forty people (10 males, 30 females), with no known auditory deficiencies, received course credit for their volunteer participation in the experiment. All were undergraduate students from the State University of New York at Stony Brook. Their first language was American English.

Materials

Twenty-four word pairs similar to "prosecutor–prosecution" were chosen (see the Appendix), all of which were four syllables long, except for "categorical" in the pair "category–categorical." The two members of each pair shared the same lexical root. Their first three syllables were identical segmentally but differed in their stress pattern. One of the words bore primary stress on the first syllable and secondary stress on the third (e.g., "prosecutor"), whereas the other word bore secondary stress on the first syllable and primary stress on the third (e.g., "prosecution"). The two types of words are hereafter referred to as *initial-primary* and *initial-secondary*, respectively. Frequency estimates based on Kučera and Francis's (1967) database revealed a slightly higher mean frequency for initial-secondary words ($M = 11.3$) than for initial-primary words [$M = 5.8$; $t(23) = 2.02$, $p < .06$]. As can be seen in Figure 1, this difference reflects the general stress pattern of four-syllable words.

In addition to the 24 test pairs, there were 70 filler pairs. As in the test pairs, the two members of each filler pair shared the segmental information of their first two syllables. However, the stress contrast between the two syllables of the filler pairs showed more variation than that in the test pairs. For instance, a stress contrast could include reduced syllables, and the contrast could lie in the first or the second syllable or in both (e.g., in the pair "affectionate"–"affectation," both the first and the second syllables show a prosodic contrast: reduced vs. secondary stressed in the first one and primary stressed vs. reduced in the second one).

Thirty practice pairs were created, using the same criteria. For recording, the 248 words, mixed into a list of 400 words, were read by four naive speakers (see the Design section). Each speaker was given a different order. The readers were not informed of the words that would be kept in the experiment, nor were they aware of the goal of the study. They were simply asked to read the words clearly and at a steady pace.

Isolating short fragments (e.g., /ˈprɑ/ or /ˌprɑ/) and long fragments (e.g., /ˈprɑsɪ/ or /ˌprɑsɪ/) from the full words (e.g., "prosecutor" and "prosecution") was accomplished with the help of a computerized speech editor. The criteria by which to locate syllable boundaries were used consistently throughout the visual editing process. All the fragments were also checked auditorily for overall perceptual quality.

Design

For a listener, deciding whether /ˈprɑsɪ/ (and, a fortiori, /ˈprɑ/) comes from "prosecutor" or "prosecution" can possibly be done either by the extraction of acoustic information within the given fragment—for example, by comparing the absolute stress value of the two syllables—or by normalization, through numerous utterances from a speaker, of a typical primary stressed syllable and a typical secondary stressed syllable. In the present design, we are more interested in the former possibility—that is, in a listener's absolute capacity to differentiate between a primary and a secondary syllable when, at best, only a reduced syllable (e.g., /sɪ/) is given for reference. To minimize the possibility that the subjects performed the task with the help of voice normalization, a given subject was exposed to the voice of two different speakers. To further control for voice independence, the two voices were either both male or both female. Presenting two same-sex voices was meant to reduce speaker normalization, while keeping the number of speakers reasonably small. Indeed, because there is potentially more misleading overlap between the stress-related acoustic features of two same-sex speakers than of two different-sex speakers (e.g., the pitch of primary stressed syllables produced by one female speaker could be roughly equivalent to the pitch of secondary stressed syllables produced by another female speaker), average normalization across the two same-sex speakers would prove relatively inefficient. On the other hand, elaborating separate normalized perceptual scales, although a more effective strategy, has been shown to be challenging for listeners (see, e.g., Strange & Gottfried, 1980; Strange, Verbrugge, Shankweiler, & Edman, 1976; van Bergem, Pols, & Koopmans-van Beinum, 1988).

The full materials were recorded four times, once by each speaker. The speakers were four students at the State University of New York at Stony Brook, whose native language was American English (the two male speakers are referred to as M_a and M_b, and the two female speakers as F_a and F_b). The listeners were randomly assigned to the

male or *female* condition (20 listeners in each). What follows is a description of the design used in both the male and the female conditions.

In the experimental phase, the subjects were presented with 188 trials. Half of them featured a short fragment (e.g., /'prɑ/ or /"prɑ/), and the other half a long fragment (e.g., /'prɑsɪ/ or /"prɑsɪ/). To be able to present both a short fragment and a long one from each of the 24 test sets to the same subject while minimizing carryover effects, the 188 trials were distributed in two blocks of 94, with a short fragment from one of the 24 sets in one block and a long fragment from the same set in the other (see Table 1). For instance, a subject presented with a short fragment of the initial-primary version of Item 1 produced by Speaker A in the first block heard the corresponding long fragment of the initial-secondary version produced by Speaker B in the second block. The combinations were rotated through the 24 pairs of items and through subjects. The 20 subjects in the male or the female talker groups were assigned to one of four subgroups so created by the Latin-square design.

The 140 filler trials were arranged as follows. Fourteen control filler trials were presented in both blocks. They were created in such a way that the two word alternatives proposed to the subjects matched the speech fragment equally well. For example, "reception" and "receptive" were proposed as word choices for the fragments /rɪ/ and /rɪ'sep/ (the fragments were recorded from a third word that shared the stress pattern of both choices). The purpose of the control trials was to identify any response tendency carried over from the first block to the second. Interblock response correlation would suggest that subjects are able to remember the occurrence of an item (or one of its versions) and to modify their response accordingly the second time they hear it. The remaining 112 filler trials were either unique to their block (56) or repeated across blocks with a different stress format (56).

Procedure

All of the stimuli were recorded in a sound-shielded booth, low-pass filtered at 4.8 kHz, digitized (12 bit A/D) at 10 kHz, and stored on the disk of a 486/100 computer. On output, the stimuli were converted to analog form (12 bit D/A) at 10 kHz, low-pass filtered at 4.8 kHz, and played over headphones at approximately 70 dB SPL.

The subjects were tested in the sound-shielded booth in groups of up to 3. Each was seated in front of a monitor and wore headphones. They were told that, on each trial, they would first see two words on the monitor and then hear the first syllable or the first two syllables of a word. The words were printed in capital letters in the middle of the screen and were 1.5 cm away from each other on the horizontal axis. The task was to decide which word on the screen best matched the spoken fragment. The subjects gave their answer by means of a response board with two buttons next to each other. If they thought that the spoken fragment was the onset of the left word on the monitor, they had to push the left button. If they thought that the spoken fragment was the onset of the right word on the monitor, they had to push the right button. The position of the items on the screen was balanced across subjects and items.

The 30 trials in the practice block and the 94 trials in each experimental block were randomized for each subject. The two words were displayed on the monitor during the entire trial duration. Five seconds after the onset of the visual presentation, the speech fragment was played. Upon speech offset, the subjects had 8 sec to hit a button. After they gave a response or after the 8-sec response window, the program cleared the screen for 2 sec and then moved on to the next trial. A 5-min break was allowed between the two blocks.

It should be mentioned that the subjects heard no entire words in the course of the experiment. The only speech presented to them consisted of one- or two-syllable fragments. Similarly, the subjects were never given feedback on their guessing performance. These two aspects, together with the dual-voice feature, minimized the possibility that the subjects could base their guesses on within-speaker syllable normalization.

Acoustic Measurements

The main acoustic parameters traditionally assumed to underpin stress perception are the F_0, duration, and amplitude of syllables (Beckman, 1986; Fry, 1958; Lehiste, 1970). Such estimates are important for the present experiment, because they may reveal which acoustic feature(s) of the stimuli the subjects relied on to perform the word-guessing task. F_0, duration, and amplitude values for the first and second syllables of the 24 pairs of test items are displayed in Table 2. Syllable F_0 was assessed by computing the average pitch period duration of the four periods in the middlemost section of the most stable part of the vowel. Amplitude values, obtained on the dB scale provided by the computerized speech editor, were calculated as the peak positive departure from the zero intensity cross-line in the nucleus of the syllable.

Analyses of variance (ANOVAs) including speaker (M_a, M_b, F_a, F_b), stress pattern of the word (initial-primary vs. initial-secondary), and position of the syllable (first syllable vs. second syllable) were conducted on the mean F_0, duration, and amplitude scores separately. For the sake of completeness, the reliability of all differences in syllable position is reported, whether it is licensed or not by higher level interactions. The overall picture that emerges from these data is that the first syllable of initial-primary words (e.g., /'prɑsɪ/) is higher pitched and longer than the second syllable in the same word (e.g., /'prɑsɪ/). This contrast is virtually absent between the first and the second syllables of initial-secondary words (e.g.,

Table 1
Subject Assignment (Within-Group Conditions Rotate Through Items)

		Block 1			Block 2		
Group	Item	Segment	Origin	Speaker	Segment	Origin	Speaker
1	1	short	primary	a	long	secondary	b
		(/'prɑ/ from "prosecutor")			(/"prɑsɪ/ from "prosecution")		
	2	short	secondary	b	long	primary	a
		(/"pre/ from "presidential")			(/'presɪ/ from "presidency")		
	3	long	primary	a	short	secondary	b
		(/'domɪ/ from "dominating")			(/"do/ from "domination")		
	4	long	secondary	b	short	primary	a
		(/"consɪ/ from "consequential")			(/'con/ from "consequently")		
2	1	short	secondary	a	long	primary	b
		(/"prɑ/ from "prosecution")			(/'prɑsɪ/ from "prosecutor")		
3	1	long	primary	a	short	secondary	b
		(/'prɑsɪ/ from "prosecutor")			(/"prɑ/ from "prosecution")		
4	1	long	secondary	a	short	primary	b
		(/"prɑsɪ/ from "prosecution")			(/'prɑ/ from "prosecutor")		

Note—Subject assignment is identical in the male speaker and female speaker groups.

Table 2
Mean F_0, Duration, and Amplitude of the First and Second Syllables of the Test Words

	Initial-Primary				Initial-Secondary		
Speaker	First Syllable	Second Syllable	Pairwise Difference (p)	Stress Pattern × Syllable Position (p)	First Syllable	Second Syllable	Pairwise Difference (p)
				F_0 (Hz)			
M_a	109	95	.001	.01	106	100	.001
M_b	129	112	.001	.001	107	112	.001
F_a	204	197	.001	.01	200	197	n.s.
F_b	209	205	n.s.	.05	199	201	n.s.
Total	163	152	.001	.001	153	153	n.s.
				Duration (msec)			
M_a	189	133	.001	.05	171	136	.05
M_b	140	127	n.s.	n.s.	132	125	n.s.
F_a	137	138	n.s.	n.s.	137	137	n.s.
F_b	179	141	.05	.05	157	140	n.s.
Total	161	135	.001	.001	149	135	.05
				Amplitude (dB)			
M_a	58.2	55.5	.001	n.s.	56.7	55.0	.001
M_b	56.4	55.2	.01	n.s.	56.4	55.0	.01
F_a	54.9	54.0	.001	n.s.	55.0	54.2	.01
F_b	56.9	54.5	.001	n.s.	56.4	54.2	.001
Total	56.6	54.8	.001	n.s.	56.1	54.6	.001

Note—"Pairwise Difference" is the significance level of the pairwise difference between the first and the second syllables for a given stress pattern. "Stress Pattern × Syllable Position" is the significance level of the interaction between stress pattern and syllable position for a given speaker. M_a and M_b are the two male speakers; F_a and F_b are the two female speakers. n.s., $p \geq .05$.

/"prɑsɪ/). With regard to amplitude, first syllables are louder than second syllables regardless of the stress pattern of the word.

Note that the average F_0 of primary stressed syllables (M = 163 Hz; e.g., /'prɑ/) is significantly higher than that of secondary stressed syllables [M = 153 Hz; e.g., /"prɑ/, $F(1,23) = 45.56, p < .001$]. The former are also longer than the latter [161 vs. 149 msec; $F(1,23) = 28.84, p < .001$]. There is no reliable difference in F_0 or duration between the second syllables of the two types of words (152 vs. 153 Hz and 135 vs. 135 msec). Finally, even though amplitude differs significantly between primary and secondary stressed syllables [56.6 vs. 56.1 dB; $F(1,23) = 8.40, p < .01$], one should be cautious in inferring that amplitude is a potentially reliable indicator of stress degree, since the second syllable is also louder in initial-primary words than in initial-secondary words [54.8 vs. 54.6 dB; $F(1,23) = 5.12, p < .05$]. The absence of interaction between stress pattern of the word and position of the syllable indicates that initial-primary words are simply globally louder than initial-secondary words.

What these numbers suggest is that, at least in the present sample, relative F_0 and duration—but not amplitude—constitute potential cues for discriminating between primary and secondary stressed syllables. That is, provided that listeners can pick up F_0 and duration differences in the range of those present in these stimuli, they may use the size of the difference between the first and the second syllables of the words to judge whether a syllable is primary or secondary stressed.

RESULTS

Guessing Performance

First, it should be mentioned that there was no correlation between the responses to the control trials in Block 1 and Block 2 [$r = -.04$, $t(558) = -0.88, p = .38$]. This result suggests that the response given to an item in Block 1 did not influence that given to another version of the same item in Block 2. Given the nature of the present design, in which each test item is presented in both blocks, independence between blocks is highly desirable.

Displayed in Table 3 are the mean percentages of correct identification of word fragments calculated for each cell generated by the design: speaker (M_a, M_b, F_a, F_b), stress pattern of the source word (initial-primary vs. initial-secondary), and fragment size (first syllable vs. first two syllables). An ANOVA performed on these data by subjects (F_1 results) and by items (F_2 results) revealed an effect of fragment size [$F_1(1,36) = 12.91, p < .001$; $F_2(1,23) = 6.39, p < .02$]. The subjects were better at guessing the source of a fragment when this fragment was long (e.g., /'prɑsɪ/ or /"prɑsɪ/) than when it was short (e.g., /'prɑ/ or /"prɑ/). Despite the higher frequency of initial-secondary-stress words than of initial-primary-stress words, both in the sample and in four-syllable-long English content words, there was no reliable bias toward responding *initial-secondary* [$F_1(1,36) = 2.77, p > .10$]. No other effect or interaction was significant, with the exception of a speaker × stress pattern interaction that was significant only by items [$F_2(3,69) = 4.09, p < .01$; $F_1(1,36) = 1.93, p = .14$]. An examination of this interaction revealed a stress pattern effect for speaker F_a, who generated better overall performance with initial-secondary than with initial-primary words [$F_1(1,9) = 11.05, p < .01$; $F_2(1,23) = 10.81, p < .005$]. No stress pattern effect was observed with any of the other speakers (all Fs < 1).

The advantage found for long fragments over short ones indicates that, despite the fragments' equal *segmental* inability to cue the correct source word (the segmental in-

Table 3
Percentages of Correct Identification of the Source Word as a Function of Speaker, Stress Pattern, and Size of the Fragment

	Initial-Primary		Initial-Secondary	
Speaker	Short Fragment (First Syllable Only)	Long Fragment (First and Second Syllables)	Short Fragment (First Syllable Only)	Long Fragment (First and Second Syllables)
M_a	49	55	53	63
M_b	62	67	53	72
F_a	46	48	67	62
F_b	51	66	52	63
Total	52	59	56	65

Note—M_a and M_b are the two male speakers, and F_a and F_b are the two female speakers.

formation in /pra/ and /prasi/ is useless for distinguishing "prosecutor" from "prosecution"), their *suprasegmental* specifications are sufficient to improve the guessing performance. If the listeners had been insensitive to primary/secondary stress differences, appending a reduced syllable to the stressed syllable in question (e.g., from /'prɑ/ to /'prɑsɪ/ or from /"prɑ/ to /"prɑsɪ/) should not have improved the guessing performances. Thus, the fragment size effect suggests that stress perception can be sharpened by providing subjects with a perceptual "yardstick" (a reduced syllable), by reference to which the degree of stress of another syllable is evaluated.

An analysis of the difference between the subject's performances in the long-fragment condition (M = 62%) and a 50% chance level to guess correctly the word's origin proved highly significant [$F_1(1,36) = 42.97$, $p < .001$; $F_2(1,23) = 39.39$, $p < .001$]. The performance in the short-fragment condition (M = 54%) also departed significantly from the chance level [$F_1(1,36) = 4.44$, $p < .05$; $F_2(1,23) = 6.71$, $p < .02$]. Thus, even though accuracy in the latter condition is low, isolated syllables might yet contain enough information to enable subjects to guess significantly better than chance whether they are primary stressed or secondary stressed. The next step is to identify some of the acoustic cues that may have guided the subjects' responses in both the short- and the long-fragment conditions.

Relationship Between Acoustic Content of the Stimuli and Subjects' Responses

In an attempt to analyze some of the possible acoustic cues that the subjects relied on to guess the origin of a word, correlation coefficients were computed between the tendency to classify a given item as initial-primary and three acoustic factors assumed to be major correlates of stress perception: fundamental frequency (F_0), duration, and amplitude (Lehiste, 1970). For example, in the short-fragment condition, if the subjects used F_0 to estimate stress degree, with high frequencies being taken as an indication of primary stress, there should be a positive correlation between the percentage of *initial-primary* responses to an item and the F_0 of this item. Similarly, in the long-fragment condition, if the subjects used the magnitude of the F_0 difference between the first and the second syllables to infer the lexical origin of the fragment, there should be a positive correlation between the percentage of *initial-primary* responses to an item and the inter-syllable F_0 difference for this item. The same rationale holds for duration and amplitude, with high values on both being associated with primary stress.

As can be seen in the left part of Table 4, the correlation results in the short-fragment condition do not offer a very coherent picture. If anything, the subjects seemed to take high F_0 and low amplitude as indications of primary stress [$r = .11$, $t(190) = 1.57$, $p = .12$, and $r = -.14$, $t(190) = -1.92$, $p < .06$, respectively]. Duration of the fragment did not influence the response [$r = .02$, $t(190) = 0.33$, $p = .74$].

To further explore the relationship between these variables, a simultaneous multiple regression was performed between the subjects' tendency to respond *initial-primary* as the dependent variable (DV) and the F_0, duration, and amplitude of the test syllable as the independent variables (IVs). Analyses were carried out, using StatView

Table 4
Standard Multiple Regression of F_0, Duration, and Amplitude on Subjects' Responses to Short Fragments (First Syllable Only)

Variable	Response r	F_0 r	Duration r	Amplitude r	β	sr^2
F_0	.11	1.00			.13	.02
Duration	.02	.12	1.00		.01	.00
Amplitude	−.14	.11	.01	1.00	−.15	.02*
					$R = .19$	
					$R^2 = .04$[a]	
					adjusted $R^2 = .02$	

Note—"Response" refers to the percentage of "initial-primary" responses across initial-primary and initial-secondary trials. [a]Unique variability = .04; shared variability = .00. *$p < .05$.

Multiple Regression. The right part of Table 4 displays the correlations between the variables, the standardized regression coefficients (β), the squared semipartial correlations (sr^2, the unique contribution of each IV to the variability of the DV), R, R^2, and adjusted R^2. The regression coefficient R (.19) showed a trend toward significance [$F(3,188) = 2.33$, $p = .07$]. Among the IVs, only amplitude contributed significantly (2%) to the variability of the DV. Neither frequency nor duration accounted for a significant fraction of the DV (2% and 0%, respectively). Thus, altogether, only 4% of the variability in the subjects' responses was predicted by knowing the scores on the three IVs.

This regression model suggests that knowing the F_0, duration, and amplitude of the test syllable (e.g., /ˈprɑ/ or /ˌprɑ/) is not sufficient to predict the subjects' responses. In addition, the negative correlation between amplitude and primary-stress judgment contradicts the literature on perceived stress, which shows that, when amplitude level has any influence on stress perception, it correlates positively with degree of perceived stress (Fry, 1958; Lehiste, 1970). Thus, in light of this regression model, it seems reasonable to conclude that the difference between the accuracy scores (54%) and the chance level (50%) reflects idiosyncratic cues in the stimuli, rather than a systematic (and adequate) exploitation of acoustic cues to stress.

The acoustic data in the long-fragment condition are more straightforward. The guessing performance in this condition revealed that the subjects were able to discriminate between primary and secondary stressed syllables better if the stressed syllable was accompanied by an unstressed reference syllable. The following analyses are designed to pinpoint the incidence of F_0, duration, and amplitude in the listeners' performance improvement. As in the short-fragment condition, the subjects' response tendency was estimated by the percentage of *initial-primary* responses for a given item. Acoustic variables were computed differently for, on the one hand, frequency and amplitude and, on the other hand, duration. For frequency and amplitude, the statistic was a difference between the values in the two syllables (i.e., $F_{0\,dif} = F_{0\,syll1} - F_{0\,syll2}$, and $amplitude_{dif} = amplitude_{syll1} - amplitude_{syll2}$), whereas, for duration, it was a ratio between the duration of the first syllable and the total duration of the first and second syllables ($duration_{ratio} = duration_{syll1}/[duration_{syll1} + duration_{syll2}]$). In all three cases, high values should be associated with primary stress.

Correlation coefficients between subjects' responses and the three acoustic parameters (left part of Table 5) reveal that the tendency to respond *initial-primary* was correlated positively with intersyllable F_0 difference [$r = .32$, $t(190) = 4.72$, $p < .0001$] and duration ratio [$r = .20$, $t(190) = 2.88$, $p < .005$]. Correlation with intersyllable amplitude difference did not reach significance [$r = .06$, $t(190) = 0.89$, $p > .30$].

A simultaneous multiple regression was performed between the subjects' tendency to respond *initial-primary* as the DV and intersyllable F_0 difference, duration ratio, and amplitude difference as the IVs. The results can be seen in the right part of Table 5. The regression coefficient R(.38) was highly significant [$F(3,188) = 10.52$, $p < .0001$]. Two IVs contributed significantly to the subjects' responses: intersyllable F_0 difference ($sr^2 = .09$) and intersyllable duration ratio ($sr^2 = .04$). Intersyllable amplitude difference did not contribute significantly to the variability of the DV ($sr^2 = .00$). The three IVs in combination contributed another 0.01% in shared variability. Altogether, 14% of the variability in the subjects' responses was predicted by knowing the intersyllable F_0, duration, and amplitude scores.

This regression model indicates that knowing the F_0, duration, and amplitude of the two syllables of the test fragments (e.g., /ˈprɑsɪ/ or /ˌprɑsɪ/) is sufficient to predict the subjects' responses (e.g., "prosecutor" vs. "prosecution") with a reliable degree of accuracy. The listeners not only used the intersyllable acoustic information to guide their perceptual judgment, but also used it appropriately. That is, high frequency and long duration were correctly associated with the occurrence of initial-primary words. In contrast, amplitude differences were not significantly relied on, an appropriate choice as well, given the lack of stress-discriminating amplitude cues in the stimuli. The present hierarchy of acoustic cues (frequency

Table 5
Standard Multiple Regression of Intersyllable F_0 Difference, Intersyllable Duration Ratio, and Intersyllable Amplitude Difference on Subjects' Responses to Long Fragments (First and Second Syllables)

Variable	Response r	F_0 Difference r	Duration Ratio r	Amplitude Difference r	β	sr^2
F_0 difference	.32‡	1.00			.30	.09***
Duration ratio	.20†	.06	1.00		.20	.04**
Amplitude difference	.06	.13	−.21†	1.00	.07	.00
					R = .38***	
					R^2 = .14[a]	
				adjusted	R^2 = .13	

Note—"Response" refers to the percentage of "initial-primary" responses across initial-primary and initial-secondary trials. [a]Unique variability = .13; shared variability = .01. **$p < .01$. ***$p < .001$.

followed by duration, itself followed by amplitude) is in accordance with most studies of stress perception (e.g., Fry, 1955, 1958; Lehiste, 1970; Morton & Jassem, 1965; Rietveld & Koopmans-van Beinum, 1987; van Heuven & Menert, 1996).

DISCUSSION

The contribution of stress to speech processing has recently become a critical object of inquiry for a number of word recognition models. However, despite the fact that linguists have devoted a great deal of attention to phonological theories that make provision for several degrees of stress (e.g., Burzio, 1994; Hayes, 1995; M. Y. Liberman & Prince, 1977; Selkirk, 1984), psycholinguists have traditionally ascribed only two levels to stress: presence of stress (full vowels) and absence of stress (reduced vowels). This dichotomous approach to stress in models of word recognition is based on two assertions. First, making fine distinctions between stress levels is not necessary to segment the vast majority of words out of the speech stream. Second, listeners cannot perceive subtle stress contrasts. In this study, I have tried to show that neither assertion is completely justified and, hence, that word recognition models could benefit from including finer stress distinctions in their processing algorithms.

The theoretical advantage of encoding more than two levels of stress was discussed in the framework of stress-based segmentation procedures. In the introduction, it was argued that the number of false detections of word boundaries could be considerably reduced if only primary stressed syllables were postulated as word onsets. Thus, even though fine stress distinctions may not be critical in distinguishing words from one another (there are only a handful of words that differ solely on primary/secondary stress distinctions), they constitute a potentially rich source of information with respect to finding word boundaries.

Furthermore, the perceptual ability to discriminate primary from secondary stressed syllables, a necessary condition for the above-mentioned hypothesis, was hereby tested in an experiment on word guessing. Subjects were presented with word fragments whose only difference was the stress degree of their first syllable (e.g., /'prɑ/ vs. /"prɑ/ or /'prɑsɪ/ vs. /"prɑsɪ/) and were asked to guess the full word from which each fragment originated in a two-alternative forced-choice task (e.g., "prosecutor"–"prosecution"). The results showed that, when the subjects were only given the critical syllable of the words (e.g., /'prɑ/ or /"prɑ/), guessing performance was very low but significantly better than chance level. However, a regression analysis revealed that some potentially useful acoustic features of the syllable (frequency, duration, and amplitude) were not used adequately, or not significantly so, by the subjects. In contrast, when the subjects were presented with longer fragments, which differed from the short fragments by the addition of a reduced syllable (e.g., /'prɑsɪ/ or /"prɑsɪ/), guessing performances were far better. A regression model revealed that the F_0 and duration of the stressed syllable, relative to those of the reduced syllable, were reliable predictors of the subjects' responses.

It should be noted that the three acoustic factors analyzed as predictors of stress perception (F_0, duration, and amplitude), although considered to be major correlates of perceived stress, might not be the only cues used by the listeners in this experiment. Pitch contour, spectral distribution, absolute pitch reference values, and so forth could also have contributed to the guessing performance. Inclusion of these in the regression models could explain an additional fraction of the variance in the results. There could also possibly be segmental influences on stress perception, with certain vowels and/or syllable structures promoting more efficient stress cues than others.[2] Explicitly manipulating these variables could be done in future experiments.

From the present results, we can conclude that listeners have the perceptual capacity to distinguish a primary from a secondary stressed syllable insofar as this syllable can be weighed against a reference syllable. Crucially, however, the reference syllable does not need to be the alternative stressed syllable with which the test syllable competes (e.g., a secondary stressed syllable if the test syllable is primary stressed, and vice versa). This finding runs counter to the idea that degrees of stress are perceptually defined only relative to one another (Cutler & Butterfield, 1992). In this experiment, guessing performances were substantially improved by the mere presence of a reduced syllable. Thus, listeners appear to be able to gauge stress relative to a single durable standard—namely, a reduced syllable.[3] How listeners evaluate the contrast between a primary or a secondary stressed syllable and such a standard is still unclear. The calculation probably entails assessing both quantitative/auditory (e.g., pitch, duration, loudness) and qualitative/phonetic (e.g., vowel quality) differences between stressed and unstressed syllables. However, as was demonstrated by this experiment, in which phonetic information was kept constant, judgments based solely on auditory differences can yield satisfactory stress perception. In any case, provided that stress calibration is completed early on in a speaker's utterance, any subsequent syllable could then be rapidly assigned its prosodic status.

One critical aspect of the present performance data is worth examining. In this study, the hypothesis that one cannot discriminate between different degrees of stress was rejected, because subjects were found to correctly distinguish primary from secondary stressed sequences 62% of the time. Although the departure from chance was statistically indisputable both by subjects and by items, the absolute performance was far from perfect. Speech engineers would no doubt be discouraged from building a speech recognition system operating with this

level of uncertainty. However, several aspects of the study should be borne in mind. First, the major issue at stake is the ability to perceive stress differences (ability as competence) and not the level of performance achievable by tapping this ability. That is, what is challenged here is the notion that we do not have the necessary competence to discriminate primary from secondary stress, a notion usually put forth to account for the nonincorporation of fine-grained stress distinctions into models of spoken word recognition. The present results succeed in showing that this hypothesis has to be rejected. Second, the degree of performance supported by this perceptual competence is likely to be a function of the quantity and quality of the information provided. The findings described here were obtained with stimuli reduced to their simplest expression. Guessing performance would presumably be higher if the sequences were presented in a richer linguistic context, which is usually the case in the machine speech recognition domain. Finally, the regression model for the long fragments indicated that relevant acoustic information (F_0 and duration) was exploited in a sensible way to deduce stress degree, whereas less relevant information (amplitude) was largely ignored. This suggests that, whenever pertinent information about stress is available, listeners tend to use it. Thus, the level of the guessing performance is contingent on the relevant cues being present in the signal, and the acoustic measurements of the stimuli revealed that some speakers gave out such cues less often than others (see Table 2).

With respect to speech segmentation, the present findings suggest that any challenge to the hypothesis that fine stress distinctions may contribute to speech segmentation cannot be made on perceptual grounds. The data show that, even with contextual information kept to a minimum, subjects were able to reliably infer a syllable's degree of stress. Likewise, the results indicate that it is not necessary to access the lexical representation of a word to assign its syllables their correct stress—how efficient would a segmentation strategy be if one of its prerequisites consisted of proper word recognition? Even though the present findings can on no account be used as a demonstration that only primary stressed syllables initiate segmentation, they show that our perceptual system can pick up the acoustic features that differentiate them from secondary stressed syllables and, hence, that listeners could, in principle, use such sensitivity to limit lexical access to primary stressed syllables.

However, the level of absolute performance (62%) should warn us that fine stress distinctions can sometimes be overlooked by listeners and that this shortcoming could presumably be reflected in the segmentation outcome. In line with this possibility, Cutler (1986) found that pairs of words like "FORbear" and "forBEAR," which have primary and secondary stress in mirror positions, behaved like lexical homophones, with words semantically related to either meaning being activated by both words. Cutler interpreted this result as an indication that fine stress distinctions were not involved in lexical access. Similarly, Luce and Cluff (1998) observed that spondees (compound words bearing stress in both syllables; e.g., "hem*lock*") can generate lexical activation on their second syllable, despite the fact that such words are typically realized with prominent stress on the first syllable. Even though spondees cannot speak to the complete range of cases for which stress-based segmentation is relevant, Luce and Cluff's results may suggest that the distinction between primary and secondary stressed syllables is not always used in words containing only stressed syllables. An alternative possibility is that, as was proposed in the introduction, stressed syllables all generate lexical activation but that they do so with a magnitude proportional to their degree of stress. Graded activation is theoretically not incompatible with Luce and Cluff's results, since lexical activation on the second syllable of the spondees (e.g., "hem*lock*") was compared with that produced by the monosyllabic version of the second syllable (e.g., "lock") and not with the activation produced by the first syllable, which would have been a more critical test. Thus, their findings do not discount the hypothesis that the initial (primary stressed) syllable was weighted more than the second one in the activation process. In fact, recent data from Vroomen and de Gelder (1997) show that replacing a word-onset primary stressed syllable with a secondary stressed syllable slows down word spotting, which suggests that primary stress is a more potent lexical *activator* than secondary stress is.

Other data also indicate that the primary/secondary stress distinction may be an important factor in the processing of longer words. Mattys and Samuel (1997) showed, in a phoneme migration experiment featuring four-syllable-long items, that the probability of misperceiving the vowel of a secondary stressed syllable (in either the first or the third position) was lower in a word than in a matched nonsense word, whereas there was no such lexical effect with the vowel of primary stressed syllables. This stress-based difference in lexical facilitation could lend support to the notion that the processing of secondary stressed syllables is influenced by the lexical information activated by other parts of the word. In contrast, primary stressed syllables would be processed more autonomously—with less assistance from the lexicon—which is consistent with the hypothesis that primary stress plays a critical role in initiating lexical access. As was described earlier, similar results were obtained by the same authors, who observed that degrading a word-late syllable slowed down the detection of an earlier syllable to a larger degree if the late syllable was primary stressed than if it was secondary stressed. The data collected in the present study show that the acoustic differences between these two types of syllables are noticeable to listeners and could, therefore, constitute effective cues for speech segmentation and lexical access.

REFERENCES

BECKMAN, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Floris.

BRANDT, J. F., RUDER, K. P., & SHIPP, I., JR. (1969). Vocal loudness and effort in continuous speech. *Journal of the Acoustical Society of America*, **46**, 1543-1548.

BURZIO, L. (1994). *Principles of English stress*. Cambridge: Cambridge University Press.

CARLSON, R., GRANDSTROM, B., LINDBLOM, B., & RAPP, K. (1973). Some timing and fundamental frequency characteristics of Swedish sentences: Data, rules, and perceptual evaluation. *Speech Transmission Laboratory Quarterly Progress Status Report*, **4**, 11-19.

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

COLE, R. A., & JAKIMIK, J. (1980b). A model of speech perception. In R. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-163). Hillsdale, NJ: Erlbaum.

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

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

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

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

ECHOLS, C. H., CROWHURST, M. J., & CHILDERS, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory & Language*, **36**, 202-225.

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.

FRY, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, **27**, 765-768.

FRY, D. B. (1958). Experiments in the perception of stress. *Language & Speech*, **1**, 126-152.

GROSJEAN, F., & GEE, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, **25**, 135-155.

HAYES, B. (1995). *Metrical stress theory*. Chicago: Chicago University Press.

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.

HERMES, D. J., & VAN GESTEL, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, **90**, 97-102.

JUSCZYK, P. W., CUTLER, A., & REDANZ, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, **64**, 675-687.

KLATT, D. H. (1980). Speech perception: A model of acoustic–phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243-288). Hillsdale, NJ: Erlbaum.

KUČERA, H., & FRANCIS, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

LIBERMAN, A. M., & STUDDERT-KENNEDY, M. (1978). Phonetic perception. In R. Held, H. W. Leibowicz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (pp. 143-178). Berlin: Springer-Verlag.

LIBERMAN, M. Y., & PRINCE, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249-336.

LIBERMAN, M. Y., & STREETER, L. A. (1978). Use of nonsense-syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America*, **63**, 231-233.

LIEBERMAN, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, **32**, 451-454.

LIEBERMAN, P. (1965). On the acoustic basis of perception of stress by linguists. *Word*, **21**, 40-54.

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

LUCE, P. A., & CLUFF, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, **60**, 484-490.

MARCUS, S. M. (1984). Recognizing speech: On the mapping from sound to word. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 151-163). Hillsdale, NJ: Erlbaum.

MATTYS, S. L., JUSCZYK, P. W., LUCE, P. A., & MORGAN, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, **38**, 465-494.

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

MATTYS, S. L., & SAMUEL, A. G. (in press). Implications of stress pattern differences in spoken word recognition. *Journal of Memory & Language*.

MCQUEEN, J. M., CUTLER, A., BRISCOE, T., & NORRIS, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language & Cognitive Processes*, **10**, 309-331.

MELTZER, R. H., MARTIN, J. G., MILLS, C. B., IMHOFF, D. L., & ZOHAR, D. (1976). Anticipatory coarticulation and reaction time to phoneme targets in spontaneous speech. *Phonetica*, **37**, 159-168.

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

MORGAN, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory & Language*, **35**, 666-688.

MORTON, J., & JASSEM, W. (1965). Acoustic correlates of stress. *Language & Speech*, **8**, 148-158.

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). Contribution of prosody to speech perception. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 75-107). New York: Wiley.

NORRIS, D., MCQUEEN, J. M., & CUTLER, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1209-1228.

PIERREHUMBERT, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, **66**, 363-369.

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

RIETVELD, A. C. M., & KOOPMANS-VAN BEINUM, F. J. (1987). Vowel reduction and stress. *Speech Communication*, **6**, 217-230.

SELKIRK, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

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.

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**, 2471-2485.

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**, 503-513.

STRANGE, W., & GOTTFRIED, T. (1980). Task variables in the study of vowel perception. *Journal of the Acoustical Society of America*, **68**, 1622-1625.

STRANGE, W., VERBRUGGE, R., SHANKWEILER, D., & EDMAN, T. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, **60**, 213-224.

VAN BERGEM, D. R., POLS, L. C. W., & KOOPMANS-VAN BEINUM, F. J. (1988). Perceptual normalization of the vowels of a man and a child in various contexts. *Speech Communication*, **7**, 1-20.

VAN HEUVEN, V. J., & MENERT, L. (1996). Why stress position bias? *Journal of the Acoustical Society of America*, **100**, 2439-2451.

VROOMEN, J., & DE GELDER, B. (1997). *Trochaic rhythm in speech segmentation*. Paper presented at the 38th Meeting of the Psychonomic Society, Philadelphia.

VROOMEN, J., TUOMAINEN, J., & DE GELDER, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory & Language*, **38**, 133-149.

WAIBEL, A. (1986). Suprasegmentals in very large vocabulary word recognition speech perceptions. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines* (pp. 159-186). New York: Academic Press.

NOTES

1. Reiterant speech is created by substituting all the syllables in a sentence or a word by a unique syllable (e.g., "ma"), thus preserving only the suprasegmental features of the original utterance. For example, the sentence "this is an utterance" becomes *'ma ma ma 'mamama* after being converted into reiterant speech (M. Y. Liberman & Streeter, 1978).

2. Analyses were carried out to measure the correlation between the quality of the vowel in the first syllable and the performance in either the short- or the long-fragment condition. None of the vowel type categorizations investigated correlated significantly with performance. Categorizations included height, backness, rounding, and so forth. Likewise, consonant type and syllable structure did not appear to modulate the performance. However, these results cannot be taken as definitive evidence that stress perception is not influenced by segmental and syllabic factors, because these were not systematically manipulated in the design and, as a result, (1) categories sometimes included only one or two tokens, and (2) vowel, consonant, and syllable structure categories were confounded.

3. The possibility that the subjects judged the stress level of a syllable by reference to a memorized token of the alternative stressed syllable presented earlier in the experiment is unlikely. First, the experiment was designed to minimize the encoding of syllable prototypes in the following ways: (1) Two voices were used in each condition, (2) the two stress-contrasted alternatives were always presented in different voices, (3) entire words were never presented, and (4) the subjects were never given any feedback on their performance. Thus, the stress level of a syllable could never be established with certainty. Second, had the subjects nevertheless been able to base their responses on memorized stressed syllable prototypes, short and long segments should have benefited equally from a comparison with these prototypes. Yet, responses were considerably less accurate with short segments.

APPENDIX
Test Items

Initial-Primary	Initial-Secondary
prosecutor	prosecution
delegating	delegation
presidency	presidential
category	categorical
consequently	consequential
navigator	navigation
vindicating	vindication
fabricating	fabrication
segregating	segregation
replicating	replication
hesitating	hesitation
agitating	agitation
celebrating	celebration
indicator	indication
calculated	calculation
generator	generation
fascinating	fascination
dominating	domination
terminating	termination
decorator	decoration
demonstrator	demonstration
cultivating	cultivation
aggravating	aggravation
ceremony	ceremonial

(Manuscript received December 22, 1997;
revision accepted for publication November 3, 1998.)