11 Prosody and Sentence Perception in English

Anne Cutler
Centre for Research on Perception and Cognition
University of Sussex

Spoken sentences are strings of segments that vary not only in spectral quality but in pitch, amplitude, and relative length. To understand what a speaker has communicated, a listener must identify the segments composing the utterance, and the order in which they occur; but a good deal of information is also encoded in the pitch, amplitude, and timing variations, i.e., the prosody of the utterance, and listeners make good use of it. Prosodic processing can be so efficient that listeners will attend to prosodic continuity at the expense of semantic continuity (Darwin, 1975). This paper addresses the question of the extent to which listeners make use of available prosodic cues, and the types of information they extract from them. All of the experimental studies cited in the three main sections below deal with the processing of English only.

A central topic of my own research has been the processing of sentence accent, or primary sentence stress, and the first section of this paper describes a series of experiments that addresses this issue. In sections 2 and 3 these studies are integrated with other work on the processing of prosody, in particular on the use of prosodic information to locate syntactic boundaries and to ascertain lexical stress patterns.

1. THE PERCEPTION OF SENTENCE ACCENT

All the studies to be reported below used the phoneme-monitoring task, a technique for studying sentence comprehension processes devised by D. J. Foss (1969). In a phoneme-monitoring experiment, subjects listen to

sentences and press a button as soon as they hear a word beginning with a specified sound. Reaction time to the target sound is significantly faster when the target-bearing word bears sentence accent—as it does for instance in (1a) and (2a) below—than when it does not (1b, 2b), and this is true both when the target-bearing word is an open class word such as a noun, verb or adjective (2) and when it is a closed class word such as a conjunction or preposition (1).

(1) target: /b/ a.   I'm not sure Shakespeare's plays are even BY Shakespeare.
             b.   I'm not sure Shakespeare's plays are EVEN by SHAKESPEARE.
(2) target: /k/ a.   Does John really want to KEEP that old van?
             b.   Does John really WANT to keep that old van?

These findings were reported by Cutler and Foss (1977). Similarly, if the target sound begins a nonsense word, response time is faster if the initial (target-bearing) syllable of the nonsense word is stressed than if it is unstressed (Shields, McHugh, & Martin, 1974).

In a subsequent experiment, it was demonstrated that the reaction time advantage of accented target bearing words is not solely due to the heightened acoustic clarity that without doubt prevails in stressed syllables in comparison with unstressed syllables. Cutler (1976) recorded sentences in two prosodic versions, one in which the target-bearing word received primary sentence accent (e.g., 3a) and one in which it did not (3b):

(3) target: /b/ a.   The couple had quarreled over a BOOK they had read.
             b.   The couple had quarreled over a book they hadn't even READ.

The target-bearing word itself was then spliced out of each version and replaced by acoustically identical copies of the same word taken from a third recording of the same sentence in which the stress level of the target word was intermediate, falling between the other two versions. This resulted in two versions of each experimental sentence, with acoustically identical target-bearing words but markedly different prosodic contour on the words preceding the target: In one case the prosody was consistent with sentence accent occurring at the location of the target, in the other case it was consistent with reduced stress at that point.

Under these circumstances the target in the "accented" position is still responded to significantly faster than the target in the "unaccented" posi-

The words in upper case bear sentence accent.

tion. Since there were no acoustic differences between the target words themselves that could account for this result, and the only difference in the preceding context lay in the prosody, it was concluded that listeners had been making use of prosodic cues that indicated where accent would occur.

The way in which the acoustically identical target words were spliced into the differing prosodic contexts in this experiment, however, allowed the interpretation that the effective component of the preceding prosody was in the tenth of a second or so immediately preceding the splice. The word-initial splices were made as close as possible to the release of the stop consonant burst (all target words began with /b/, /d/ or /k/). The closure before the burst of a stop consonant, however, tends to be longer in stops preceding stressed vowels than in stops preceding unstressed vowels, and in fact in the Cutler (1976) materials the stop closures were consistently longer before targets in "accented" contexts than before those in "unaccented" contexts. This could have allowed subjects more processing time in the "accented" case, which in turn could have produced faster response times. (This argument is directly analogous to the explanation given by Mehler, Segui, & Carey, 1978, for the fact that reaction time to phoneme targets is faster when the word before the target is polysyllabic rather than monosyllabic.)

In order to investigate this possibility, the preceding experiment was replicated with position of the word-initial splice and hence duration of the stop closure explicitly manipulated: Splicing at the stop burst was compared with splicing at the onset of the pre-burst closure (Cutler & Darwin, 1981). The conditions of the previous experiment were repeated except that each experimental sentence occurred in four versions, resulting from the comparison of the two accent conditions ("accented" and "unaccented") and the two splicing position conditions (splicing at the burst and splicing at the closure onset). In the burst-splice condition the closure duration of the target stop was longer in the "accented" than in the "unaccented" versions of the sentences, whereas in the closure-splice condition both versions had stop closure durations of equal length. Again, the targets in "accented" position were responded to significantly faster than the targets in "unaccented" position, but the splicing manipulation had no effect: The reaction time advantage for targets in accented positions was equally strong in both the burst-splice and closure-splice conditions. In other words, the duration of closure of the target stop consonant does not have an effect on response time. This result allows us to reaffirm that the reaction time difference found in the earlier study was indeed due to differences in the prosodic contour preceding the target word. Listeners were using cues in the preceding

---

[2] I am grateful to R. Diehl for suggesting this possibility.

prosody that pointed to the locations of primary sentence stress; in effect, they were actively looking for accented words.

Why should listeners want to do this? A suggested answer was provided by Cutler and Fodor (1979), who pointed out that the location of primary sentence stress usually corresponds to the most informative part of a sentence, or the semantic focus. Thus when a listener determines where the primary sentence accent will fall, he has found out which part of the sentence will be semantically most central. Cutler and Fodor provided support for this interpretation by demonstrating that focused words enjoy a phoneme-monitoring response time advantage analogous to that of accented words. They presented listeners with sentences like (4):

(4) target: /k/ (or /d/)   The house with the carport must belong to the doctor's widow.

The sentence's focus was determined by preceding it with a question about some aspect of the content. Each sentence had two alternative target positions—in (4), the /k/ on "carport" and the /d/ on "doctor's"—and two alternative questions, one of which focused on the part of the sentence in which the first target occurred, while the other focused on the part where the second target occurred. For (4), the two questions were (5a) and (5b):

(5) a.   Which house must belong to the widow?
    b.   Which widow must the house belong to?

Although the sentences remained acoustically invariant—i.e., all subjects heard exactly the same recording of each sentence irrespective of which preceding question they heard or which target they were listening for—targets in earlier position produced faster responses if the preceding question had focused on the early part of the sentence, whereas targets in later position produced faster responses when the preceding question had focused on the later part of the sentence. That is, varying the position of semantic focus affects response time in much the same way as varying the position of sentence accent.

It is not clear, however, that using prosody to locate sentence accent and using semantic information within the sentence to locate sentence focus are merely two alternative procedures for accomplishing exactly the same end, one or other (but not both) to be pursued according to the information available in the utterance. Some relevant evidence was provided by a further experiment in which semantic focus and sentence accent were varied independently. In this study, carried out at the University of Sussex as an undergraduate project by Louise Lee Seng, sentences like (6) were prepared in two prosodic versions, one in which sentence accent fell on the target-bearing word and one in which it fell elsewhere. The sentences were preceded by one of two alternative questions, one of which focused on the part of

the sentence where the target occurred, while the other focused on some other part of the sentence.

(6) target: /b/   - Where did he drive yesterday?
                     - Who did he meet yesterday?
              a. Yesterday he drove to the BEACH and met his friend there.
              b. Yesterday he drove the the beach and met his FRIEND there.

The effects of semantic focus and sentence accent in this experiment turned out to be significant, equally strong, and additive, in that when the target-bearing word was both accented and focused, its reaction time advantage in comparison with the unaccented unfocused case was twice as large as when it was either accented, or focused, but not both. Note that in this experiment the sentences were not spliced; the acoustic advantages of accented over unaccented words were present and available to the listener. In the Cutler (1976) experiment the unspliced original sentences showed an accent effect more than twice as large as the effect in the spliced sentences; i.e., the effect due to prediction of accent location from preceding prosody was overlaid by a further effect due to the heightened acoustic clarity of the accented words. This may also have been the case in this experiment. Nevertheless, the RT advantage of the accented but unfocused condition—in which cues to accent were present both in the preceding prosody and in the acoustic form of the target word itself—in comparison with the unaccented unfocused condition was only half as large (46 msec) as that of the condition in which the target was both accented and focused (105 msec; focus alone produced a 50 msec RT advantage). Thus it does appear that the focus effect is separable from any accent effect.

   Clearly, the case in which accent and focus coincide is the norm in English utterances; so that it is possible that when the target-bearing word was accented but not focused, or focused but not accented, the sentences were perceived to be abnormal and an inhibitory effect thus reduced the reaction time advantage. Of course it is also possible that the reaction time advantage of accented and focused words respectively represent entirely independent phenomena that just happens to coincide in most utterances. This, though, makes no capital of the fact that the effects of focus and accent on phoneme-monitoring response time are identical in kind and remarkably similar in magnitude. Perhaps an appropriate interpretation is that search for primary sentence stress on the basis of prosodic information and search for semantic focus on the basis of non-prosodic information are separate strategies proceeding in parallel but directed at the same end. In normal sentence understanding both strategies are invoked in order that the semantically most central portions of the incoming message may be located

as rapidly as possible.

So far, by using the general term "prosody," we have begged the question of exactly what information in the acoustic signal is sufficient, or necessary, to indicate the location of sentence accent. An attempt to investigate this question more closely was undertaken by Cutler and Darwin (1981). Of the three major dimensions of prosodic variation, pitch, amplitude, and timing, the one that seemed an obvious candidate for initial study was pitch. Lea (1974), for example, has claimed that fundamental frequency variation is the most useful cue for the automatic location of sentence accent; Faure, Hirst, and Chafcouloff (1980) have reported that judgments of compound stress (e.g., "blackbird" versus "black bird") cannot be reliably made in the absence of pitch information. Thus it appeared that if any prosodic dimension were to prove a necessary source of information from which to locate sentence accent, it would most likely be pitch. Moreover, pitch is the only prosodic dimension that lends itself readily to simple test—its contribution can be gauged by eliminating pitch variation entirely, i.e., by monotonizing the sentences. Manipulation of duration and amplitude is far more difficult, and cannot simply be achieved by eliminating variation along the dimension in question, since amplitude variations are vital for many segmental distinctions, whereas uniform timing requires an a priori decision as to the relevant units of speech (phonemes? syllables? inter-stress intervals?) that should be rendered uniform in length.

Accordingly, Cutler and Darwin tested whether the absence of pitch information would alter the effectiveness of preceding prosodic cues to sentence accent location. The experiment consisted essentially of a replication of the Cutler (1976) study, with the added variable of presence or absence of pitch information. The spoken sentences were coded into LPC parameters and two versions were synthesized, one with the original pitch contour and one with a monotone pitch; the amplitude envelope of the speech was the same in each version. Thus each experimental sentence occurred in four versions, resulting from the comparison of the two accent conditions ("accented" versus "unaccented") and the two pitch conditions (intact versus monotonized). The target-bearing word was acoustically identical (monotone, neutral-stress) in all four versions of each sentence. Once again the targets in the "accented" position were responded to more rapidly than the targets in "unaccented" position, but surprisingly, the absence of pitch information made absolutely no difference to the magnitude of the accent effect. Thus it would appear that pitch variation is *not* a necessary component of the prosodic variation that enables listeners to determine in advance the location of sentence accent.

The question remains open whether either durational or amplitude variation is necessary, or whether any prosodic information is sufficient. There is

a certain amount of evidence that the information contained in durational variation may be particularly important. Phoneme-monitoring RT is extremely sensitive to manipulations of sentence rhythm. Meltzer et al. (1976) showed that minor temporal displacements of phoneme targets from their originally uttered positions (by as little as 50 msec) resulted in an increase in response time. An analogous conclusion may be drawn from an experiment of my own in which a short interval of silence was spliced into a sentence to alter the rhythm. This experiment was based on an earlier study by Martin (1970), who reported that splicing a silent interval into a sentence led to a shift in the locations at which subjects reported perceiving stress. To test whether this effect carried over into the phoneme-monitoring response advantage for sentence stress, sentences were constructed with two alternative target-bearing words, one immediately after the other, and spoken rhythmically and in such a way that stress fell on the second of the two possible target words. Thus in (7) the target could be the /d/ of "dead" or the /k/ of "cat," but stress in the original utterance fell on "cat" and not on "dead":

(7)   My friend was upset by the sight of a dead cat on the road.

A second version of each sentence was prepared in which a silent interval, equivalent to half of the rhythmic inter-stress interval, was spliced into the sentence immediately before the first of the two potential target words. This should have the effect of moving the *subjective* location of stress, i.e., the rhythmically determined point at which stress ought to occur, back to the first target word even though the second target word was the objective location of stress. For each sentence half the subjects listened for one target, and half for the other, and within each group, half heard the sentence with the silent interval and half the version without. Targets in the second position—/k/ in (7)—which were stressed in the original utterance, produced faster RTs in the original version than in the version with the rhythm-altering interval of silence. Targets in the first position—/d/ in (7)—which became the subjective location of stress once the silent interval had been added, produced faster responses in the version with the silent interval than in the version without. In other words, manipulation of the timing relationships within a sentence can have a strong effect on phoneme-monitoring response time; by implication, this effect is mediated by the effect of durational manipulations on the location of perceived stress.

It would appear, therefore, that durational variations may contain sufficient information to indicate where stressed syllables will occur. Note, however, that not all stressed syllables are accented syllables. Accent gives Prominence to one word or phrase at the expense of the rest of the sentence, and only occurs once or at most a very few times in each sentence. But in English, which is a so-called stress-timed language, there is a clear dif-

ference between stressed and unstressed syllables in general. Polysyllabic words always have one syllable marked for higher stress than the others. Within an utterance there is a rhythmic arrangement of the syllables by which, ceteris paribus, only those syllables marked for stress receive their full durational value, while unstressed syllables tend to be shortened. Sentence accent is applied only to syllables which are already marked for stress, and is in a sense overlaid on the rhythmic stress value, so that accented stressed syllables may actually be assigned even greater durational weight than non-accented stressed syllables. Thus following cues to the location of rhythmic stress beats would not be a very effective strategy for locating sentence accent, since many words are rhythmically stressed but not accented. Moreover, the recordings of the sentences used in the series of splicing experiments which began with Cutler (1976) were submitted to a panel of judges experienced in the analysis of English sentence rhythm, who marked the locations of rhythmic stress beats and of sentence accent in each sentence. The judges were in agreement that the target-bearing words were rhythmically stressed even when they were not in accented position. That is to say, rhythmic stress beats fell on both "accented" and "unaccented" target words; so if anticipation of rhythmic stress beats alone were responsible for the facilitation of response time, there should have been no difference between the two versions of each sentence. There was a difference, however; thus discussion of durational cues to the location of sentence accent must invoke not the rhythmic arrangement of stressed and unstressed syllables alone, but additional cues to the location of sentence accent.

The experiment by Shields et al. cited previously provides some further relevant evidence because it included a control condition in which the nonsense word targets from the experimental sentences were embedded in lists of other nonsense words. Under these conditions, with all the lexical stress information present but accentual variation absent, the phoneme-monitoring response time advantage for stressed-syllable over unstressed-syllable targets disappeared.

These findings make it appear likely that durational information should be of use in the location of sentence accent. Nevertheless, we are still left with the earlier report by Lea (1974) that pitch variation provides a sufficient cue to accent position. Furthermore, Huss (1975) found that fundamental frequency was a powerful cue to stressed syllable location in *accented* (but not in unaccented) words. Thus, although pitch is not a necessary source of information about the position of accent, it may well be a sufficient cue. The question of necessary versus sufficient information is also the theme of the next two sections, which consider the usefulness of the various sources of prosodic variation for making syntactic and lexical decisions during sentence processing.

2. PROSODIC CUES TO THE LOCATION OF
SYNTACTIC BOUNDARIES

There has been a very large amount of work on this topic in recent years, and among the findings are two suggesting that some caution should be exercised before results from this work are generalized to the perception of natural speech. The natural way to study perception of syntactic boundaries is to use syntactically ambiguous sentences, whether simple bracketing ambiguities such as (8) and (9) or more complex ambiguities such as (10) or (11):

   (8)   Add sage and thyme or oregano for flavor.
   (9)   He sells used Peugeots, Motobecanes, and Raleighs.
  (10)   These days few people know how good bread tastes.
  (11)   Smoking cigarettes can be offensive.

But the presence of disambiguating prosodic cues to syntax in natural speech may depend crucially on whether the speaker is aware of the ambiguity. Lehiste (1973) recorded speakers reading ambiguous sentences and then played the recordings to listeners who were asked to pick the interpretation intended by the speaker. They were fairly successful at doing so when the speaker had been consciously trying to disambiguate the sentence, but much less successful when the speaker had been unaware of the ambiguity. Exactly the same result is found with strings that are ambiguous between an idiomatic and a literal reading, such as (12) (van Lancker & Canter, 1981):

  (12)   They have all gone on the wagon.

Moreover, prosodic cues often do not suffice to disambiguate when there exists a strong frequency bias in favor of one or other reading of an ambiguous sentence (Wales & Toner, 1979); again, the same is true of idiom-literal ambiguities in which there exists a general bias towards idiomatic readings (van Lancker & Canter, 1981).

   These reservations should be borne in mind when considering the studies of prosodic cues to syntactic boundaries; most of these studies, however, have dealt with simple bracketing ambiguities like (8) and (9), which Lehiste (1973) found to be the most easily disambiguated kind of syntactic ambiguity and the most likely to be disambiguated even when the speaker was not consciously trying.

   On the one hand, we find that durational variation forms a very effective cue to syntactic boundary location. Lehiste, Olive, and Streeter (1976) reported that lengthening the foot (inter-stress interval) containing a syntactic boundary was a sufficient indicator of the presence of the boundary. To

a certain extent boundary information may be retrieved from pause dura-
tion alone (O'Malley, Kloker, & Dara-Abrams 1973). But investigation of
the exact nature of the lengthening cue by Scott (1982) has shown that the
effects both of pause and of phrase-final lengthening are primarily a reflec-
tion of their effect in lengthening the foot containing the boundary, relative
to the lengths of other feet in the utterance. Compare this result with that of
Huggins (1972), who found that listeners' perceptions of segment length
depended crucially on the effect that variations in segment duration had on
relative foot duration. Klatt and Cooper (1975) found further evidence that
listeners' expectations of segment length depend on a segment's position in
the syntactic structure of the utterance.

Thus we see that durational information can be a sufficient cue to phrase
boundary location. But on the other hand so can pitch. The location of a
fall in fundamental frequency is a highly efficient way of automatically
detecting a syntactic boundary (Lea, 1973); Collier and 't Hart (1975) also
found that listeners judge non-final falls in fundamental frequency to in-
dicate the presence of a boundary.

When listeners have the option of using a number of cues—as they
presumably do in natural speech—then the effects of pitch and duration
sum rather than interact; both of these cues are more informative than cues
provided by amplitude variations, however (Streeter, 1978). Nakatani and
Schaffer (1978), on the other hand, found duration to be a stronger cue to
word boundary location than either pitch or amplitude.

Therefore it must be concluded that listeners are very versatile in their use
of prosodic cues to syntactic structure. Either pitch or durational informa-
tion will suffice. Scholes (1971) has claimed that amplitude information
may also be relied upon. There is no indication in any of the syntactic
studies, however, that any particular prosodic cue is *necessary*.

3. PROSODIC CUES TO LEXICAL STRESS

In languages, like English, that have variable lexical stress, this information
is also carried by the prosody. Lexical stress patterns strongly influence
word perception in English—when words are misheard, stress pattern is
maintained (Garnes & Bond, 1975), and erroneously stressed words may be
perceived as other words with the (erroneous) stress pattern, in defiance of
segmental information (Bansal, 1966, cited in Huggins, 1972). False
recognition of nonsense syllables can be precipitated by stress pattern
similarity (Robinson, 1977).

Again, a great deal of research effort has been expended on investigation
of prosodic cues to lexical stress, and again, all kinds of prosodic informa-
tion have been found to be effective. Fry (1955, 1958) investigated the

perception of lexical stress in synthesized isolated words and found that both duration (1955) and pitch movement (1958) were strong and effective cues, whereas amplitude information was less useful. Similarly, Nakatani and Aston (1978) reported that duration was the best cue to perceived stress in nonsense disyllables, with fundamental frequency a less effective cue, and intensity information almost useless. Morton and Jassem (1965), however, found that while pitch movement was a better cue than either duration or intensity, the latter were about equally effective. Lieberman (1960) found effects of both fundamental frequency and intensity.

More recent work has concentrated on combinations of cues. Lea (1977) found the best algorithm for automatic location of lexical stress to be one based on a combination of pitch and intensity information. Cheung, Holden, and Minifie (1977), however, found their listeners' judgments of lexical stress to be based mainly on fundamental frequency variation. Huss (1975) found that both duration and fundamental frequency could act as indicators of syllable stress; Gaitenby (1975) found that a method of detecting stressed syllables that used all three sources of information—pitch, amplitude, and duration—was highly effective. Similarly, Isenberg and Gay (1978) found that all three parameters were sufficient cues for the perception of syllable stress; Gay (1978) has since argued that stress perception in natural speech most likely involves the use of multiple cues. Indeed, M. Smith (personal communication) has determined that listeners' perception of stress in nonsense disyllables is not fully accounted for by appealing to any first-order physical information alone, but appears to involve the use of very complex interactions of information.

It appears that for the detection of lexical stress, whatever may be the optimal combination of cues, just about any prosodic information is *sufficient*.

4. CONCLUSION

Given the findings of studies on prosodic cues to lexical stress and to syntactic structure, it does not appear surprising that pitch contour information turns out not to be a necessary cue to the location of sentence accent. The weight of the evidence indicates that with respect to prosodic processing the English speaker's sentence perception device is extremely flexible; it does not, for example, insist on durational cues to phrase boundaries and pitch cues to accent, or vice versa, but accepts and makes use of whatever information is available. Indeed, the fact that in some cases a particular prosodic dimension A has been reported to be more informative than another dimension B may reflect the fact that with certain materials or under certain experimental conditions A may have more scope for variability than B—i.e.,

there may be objectively more information available in the signal of A than of B, rather than a listener preference for A information over B.

Of course, it is also entirely possible that under natural sentence perception conditions it may be a useful strategy to use different dimensions of the prosody as sources of different information. All the experimental studies tell us is that no single prosodic cue appears to be necessary for any of the three tasks we have considered; identification of lexical stress pattern, location of accent, and location of syntactic boundaries. (Note that these are not the only uses of prosody—cues to meaning, for instance, can be extracted from particular pitch movements: Studdert-Kennedy & Hadding 1973; Mulac & Nash 1977.)

But versatility in prosodic processing should not be unexpected if one considers that a prosodic component is likely to be integral to the human language processor. Languages differ widely in the type of information they encode in the prosody (tone languages and lexical stress languages encode lexical information that other languages don't, for instance) and in the prosodic dimensions they employ (English and other "stress-timed" languages make more varied use of durational information than do French and other "syllable-timed" languages, for example). Thus if any of the work cited had proven amplitude, timing, or pitch contour to be a *necessary* cue to detection of syllable stress, sentence accent or syntactic boundaries, we would have to reckon with the possibility that it would be a necessary cue only in the perception of English, i.e., a language-specific criterion.

There is a good deal of evidence that the nature of prosodic processing differs with the language being processed (although we have only dealt with English here, Lehiste, 1970, is a source of data from languages other than English). For example, the same speakers can be shown to be making use of different cues to the same information in English and Spanish (Hutchinson, 1974). Similarly, when the same nonsense syllable stimuli were presented to English (Morton & Jassem, 1965) and Polish listeners (Jassem, Morton, & Steffen-Batog, 1968), the relative effectiveness of cues to lexical stress, in particular of durational variation, differed across the two groups. A recent parallel investigation by Berinstein (1979) compared the effect of durational variation on the perception of stress in nonsense syllable strings presented to speakers of English and of K'ekchi (a Mayan language with fixed final stress). Berinstein found that whereas duration was an effective cue to stress for the English speakers, the K'ekchi speakers tended to perceive final stress irrespective of the durational pattern of the strings. It was shown, however, that this lack of durational effect for the K'ekchi speakers was not simply due to the presence of fixed stress patterns in their language, but was of more complex origin; speakers of Cakchiquel, another Mayan language that also has fixed final stress, but differs from K'ekchi in that K'ekchi has phonemic vowel length whereas Cakchiquel does not, do find durational

variation an effective cue to lexical stress. Thus durational information is not used as a cue to stress by K'ekchi speakers because, it seems, they are already using it as a cue to vowel length. However only pre-emption of this kind appears to rule out use of a particular kind of prosodic variation as a possible source of information on which to base stress, accent or syntactic grouping decisions in any language for which such decisions have to be made.

This is not to say that prosodic characteristics of a given language do not predispose its speakers' prosodic perception. It is not surprising to find that Berinstein's Cakchiquel speakers, though they could make judgments of non-final stress on the basis of durational information, required more durational difference between stressed and unstressed syllables to make this judgment than did English speakers; Cakchiquel has final stress whereas English has variable stress (with a slight bias towards initial stress). Similarly, when duration was equal, K'ekchi speakers judged the strings to have final stress whereas English speakers judged them to have initial stress.

Interestingly, it is not clear that this language-specific effect carries over into the perception of the grouping and relative salience of non-speech sounds. Jakobson, Fant, and Halle (1952) claimed that sequences of knocks are grouped differently by Czech, Polish, and French speakers on the basis of word stress regularities in their respective languages, but Bell (1977) reports an experiment in which the perceived grouping and relative prominence of a sequence of tones heard by speakers of English, Bengali, French, and Polish showed no correspondence with the stress patterns characteristic of the languages in question. Similarly, one cannot invoke differences between English and German, which are accentually similar, to account for the fact that an identical vehicle siren sequence is perceived by English-speaking children as a high-low alternation (doo-dah, doo-dah) but by German-speaking children as a sequence of low-high tones (ta-tu, ta-tu). The predisposing effects of linguistic structure on prosodic perception may be confined to the pitch, amplitude, and duration of *speech*.

The conclusion that we draw from the available evidence, then, is that the human language processor embodies a very versatile capacity to process prosodic information. Characteristics of individual languages may determine that only a subset of the prosodic information available is normally used; but where there are multiple prosodic cues to stress, to accent or to syntax, we are *capable* of using them all.

ACKNOWLEDGMENTS

Darwin, Steve Isard, Christopher Longuet-Higgins, and Donia Scott, for many fruitful discussions of the issues involved in the perception of prosody, and to the first two in particular for helpful criticism of the present paper. They are of course not responsible for its defects.

REFERENCES

Bansal, R. K. *The intelligibility of Indian English: Measurements of the intelligibility of connected speech, and sentence and word material, presented to listeners of different nationalities.* Ph.D. thesis, London University, 1966.

Bell, A. Accent placement and perception of prominence in rhythmic structures. In L. M. Hyman (Ed.), *Studies in stress and accent.* Southern California Occasional Papers in Linguistics 4, 1977.

Berinstein, A. *A cross-linguistic study on the perception and production of stress.* UCLA Working Papers in Phonetics 47, 1979.

Cheung, J. Y., Holden, A. D. C, & Minifie, F. D. Computer recognition of linguistic stress patterns in connected speech. *IEEE Transactions on Acoustics, Speech <Sc Signal Processing,* 1977, ASSP-25, 252-256.

Collier, R., & 't Hart, J. The role of intonation in speech perception. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception.* Heidelberg: Springer, 1975.

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

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

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

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

Darwin, C. J. On the dynamic use of prosody in speech perception. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception.* Heidelberg: Springer, 1975.

Faure, G., Hirst, D., & Chafcouloff, M. Rhythm in English: Isochronism, pitch and perceived stress. In L. Waugh & C. H. van Schooneveld (Eds.), *The melody of language.* Baltimore: University Park Press, 1980.

Foss, D. J. Decision processes during sentence comprehension: Effects of lexical item difficulty and position on decision times. *Journal of Verbal Learning <£ Verbal Behavior,* 1969, *8,* 457-462.

Fry, D. B. Duration & intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America,* 1955, *35,* 765-769.

Fry, D. B. Experiments in the perception of stress. *Language & Speech,* 1958, A 120-152.

Gaitenby, J. H. Stress and the elastic syllable: An acoustic method for delineating lexical stress patterns in connected speech. *Haskins Laboratories: Status Report on Speech Research,* 1975, SR-41, 137-152.

Games, S., & Bond, Z. S. Slips of the ear: Errors in perception of casual speech. *Papers from the Eleventh Regional Meeting, Chicago Linguistic Society,* 1975.

Gay, T. Physiological and acoustic correlates of perceived stress. *Language <S Speech,* 1978, *21,* 347-353.

Huggins, A. W. F. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America,* 1972, *51,* 1279-1290.

Huss, V. Neutralisierung englischer Akzentunterschiede in der Nachkontur. *Phonetka, 1915,32,* 278-291.

Hutchinson, S. J. P. *The learning of English suprasegmental rules for stress and final syllables by Spanish speakers.* M.A. Thesis, University of Texas, 1974.

Isenberg, D., & Gay, T. Acoustic correlates of perceived stress in an isolated synthetic **disyllabic *Journal of the Acoustical Society of America,* 1978, *64,* S21.**

Jakobson, R., Fant, G., & Halle, M. *Preliminaries to speech analysis.* Cambridge, MA: MIT Press, 1952.

Jassem, W., Morton, J., & Steffen-Batog, M. The perception of stress in synthetic speech-like stimuli by Polish listeners. *Speech Analysis and Synthesis,* 1968, *1,* 289-308.

Klatt, D. H., & Cooper, W. E. Perception of segment duration in sentence context. In A. **Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception.* Heidelberg:** Springer, 1975.

Lea, W. A. *Prosodic aids to speech recognition: II. Syntactic segmentation & stressed syllable location.* Sperry Univac Technical Report No. PX 10232, 1973.

Lea, W. A. *Prosodic aids to speech recognition: V. A summary of results to date.* Sperry Univac Technical Report No. PX 11087, 1974.

Lea, W. A. Acoustic correlates of stress & juncture. In L. M. Hyman (Eds.), *Studies in stress and accent.* Southern California Occasional Papers in Linguistics 4, 1977.

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

Lehiste, I. Phonetic disambiguation of syntactic ambiguity. *Glossa,* 1973, 7, 107-122.

Lehiste, I., Olive, J., & Streeter, L. A. The role of duration in disambiguating syntactically **ambiguous sentences. *Journal of the Acoustical Society of America,* 1976, *60,* 1199-1202.**

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

Martin, J. G. Rhythm-induced judgments of word stress in sentences. *Journal of Verbal Learning & Verbal Behavior,* 1970, *9,* 627-633.

Mehler, J., Segui, J., & Carey, P. Tails of words: Monitoring ambiguity. *Journal of Verbal Learning & Verbal Behavior,* 1978, *17,* 29-35.

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

Morton, J., & Jassem, W. Acoustic correlates of stress. *Language & Speech,* 1965, *8,* 159-181.

Mulac, A., & Nash, R. Effects of intonation pattern of synthesized and natural speech upon listener resolution of semantic ambiguity. *Language & Speech,* 1977, *20,* 274-279.

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

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

O'Malley, M. H., Kloker, D. R., & Dara-Abrams, B. Recovering parentheses from spoken **algebraic expressions. *IEEE Transactions on Audio & Electroacoustics,* 1973, AU-21,** 217-220.

Robinson, G. M. Rhythmic organization in speech processing. *Journal of Experimental Psychology: Human Perception <£ Performance,* 1977, *3,* 83-91.

Scholes, R. J. On the spoken disambiguation of superficially ambiguous sentences. *Language & Speech,* 1971, *14,* 1-11.

Scott, D. R. Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America,* 1982, 77 (in press).

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

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

Studdert-Kennedy, M., & Hadding, K. Auditory and linguistic processes in the perception of intonation contours. *Language & Speech,* 1973, *16,* 293-313.

van Lancker, D., & Canter, G. J. Idiomatic versus literal interpretations of ditropically ambiguous sentences. *Journal of Speech and Hearing Research,* **1981,** *24,* **64-69.**

Wales, R. J., & Toner, H. Intonation and ambiguity. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.