High-Stress Words are Easier to Perceive than Low-Stress Words, even when they are Equally Stressed

Anne Cutler

In general, the greater the degree of stress assigned to a word in a sentence, the longer its vowel (or the vowel of its stressed syllables) will be, the higher will be the relative pitch level of the word, and, to a certain extant, the greater will be the peak amplitude of its stressed syllable. Words assigned little relative stress in a sentence, on the other hand, tend to have shorter duration, lower relative pitch, and less relative amplitude; further, the vowels in unstressed words reduce to /ə/. It is reasonable to assume that words that are longer, louder and higher-pitched will be somewhat easier to identify than words that are shorter, softer and lower-pitched and contain reduced vowels; indeed, several experiments (Lieberman 1963, Klatt and Stevens 1972, Lea, Medress and Skinner 1973) have shown this to be so.

Previous research has shown that word stress level is a relevant factor in immediate processing difficulty during sentence comprehension. Cutler and Foss (1973) presented subjects with sentences and asked them to listen for a specified word-initial phoneme and to press a button when this target was perceived - longer reaction times (RTs) in this phoneme-monitoring task (Foss 1969) are assumed to reflect greater processing difficulty. They found that stressed words produce significantly faster RTs than do unstressed words.

If this difference is due merely to the effects of heightened intelligibility; that is, to the acoustic differences between stressed and unstressed words, no important implications follow for a model of the sentence comprehension process. Other research indicates, however, that some further dimension, beyond acoustic factors, is involved. Shields, McHugh and Martin (1974), using nonsense words embedded in sentences, found similar phoneme-monitoring RT differences between targets on stressed and on unstressed syllables to those found by Cutler and Foss. When they presented the same nonsense words as a list rather than embedded in sentences, however, there was no significant difference in RT between stressed and unstressed target-bearing syllables. Shields et al. concluded from this that the difference found in the sentences was not due simply to difference in the acoustic waveform of the stressed versus unstressed syllables; they hypothesized that the listener was using sentence rhythm to predict the location of upcoming accents and direct attention towards them.

If it is indeed the case that the sentence processing mechanism utilizes suprasegmental cues to determine the location of stressed items before they actually occur, a model of the sentence comprehension process must include such an operation. Unfortunately, however, Shields et al.'s results provide only slim evidence on which to base such a claim; the processing of a string of nonsense words, as in their control sequences, may be so foreign a task to subjects that the advantage of greater intelligibility of stressed syllables is simply masked by an overall rise in level of difficulty. (Support for this claim is found in the fact that the mean RTs in their control sequences were considerably longer than the mean RTs to the same targets in the sentences.)

The critical test of whether the RT advantage of stressed words is due simply to acoustic differences, or whether some further dimension is involved, must obviously use sentences rather nonsense strings, and preferably words rather than nonsense items as target-bearers. For instance, if the same (acoustically identical) word were shown to elicit different RTs depending on the supra-

segmental context in which it occurred, the RT difference could not be due to
acoustic variation in the word itself. But is it possible to produce two sent-
ences which differ in suprasegmental contour and yet each contain a given word,
acoustically identical in each occurrence? Yes, by splicing the tape. In the
experiment to be described here, two copies of one recording of a word - i.e.
two acoustically identical sequences - were spliced into two different contexts.
These contexts consisted of the same words in each case, spoken however with
two different intonation patterns. One pattern, intact, assigned high stress to
the target-bearing item, the other low stress. Once the original target-bearing
word had been removed from each context and the acoustically identical replace-
ments inserted, the one context predicted that the target-bearing item would bear
high stress, while the other predicted that it would bear low stress. If diff-
erences in intelligibility were soley responsible for the previous RT advantage
of stressed words, no difference should be found betwe.n RTs to the target in
the two different contexts. If, however, part of sentence processing is the
prediction of upcoming stress locations, and the RT advantage is due at least
in part to this, then the target word should produce faster RTs when it is em-
bedded in the context which predicts that it will bear high stress than when
it is in the context which predicts its stress level will be low.

The materials used in the experiment were as follows: twenty unrelated
sentences were recorded in three versions. In one version, the target-bearing
word was heavily stressed. In the second version, that word received very re-
duced stress. In the third version neutral, or intermediate, stress was assigned
to the target word. In order to make the intonation contours sound natural,
the three versions had different endings; however, the point at which the sent-
ences diverged was beyond the occurrence of the target. The target-bearing word
was in each case a monosyllabic noun beginning with one of the three phonemes
used as targets - /b/, /d/, /k/ - and it occurred more than five syllables after
the beginning of the sentence, and not at the end of the sentence.

An example sentence is the following (phoneme target /d/):

High stress on target: She managed to remove the dirt from the rug, but not the
berry stains.

Low stress on target: She managed to remove the dirt from the rug, but not
from their clothes.

Neutral (control) version: She managed to remove the dirt from the rug.

The relative increase and decrease in target item stress of the experimental
versions in comparison with the control version has been obtained, as can be
seen from the different endings, by manipulation of what is commonly called
contrastive stress.

The target words were spliced out of all three versions of each sentence.
The high- and low-stress target items were discarded, and a copy of the target
item from the control version of each sentence was spliced into their place in
the first and second versions. The experimental sentences thus consisted of
two versions of each sentence - one in which the intonation contour predicted
high stress on the target-bearing word, and one in which the contour predicted
low stress on that word. The target word itself was, however, intermediate in
stress, and, moreover, identical in each version.

Two tapes were constructed, each containing one version of each of the
twenty experimental sentences plus forty filler items, twenty of these without

occurrence of the target. Predicted-high versus predicted-low stress was counter-balanced across the two tapes for the experimental items. Two further tapes were compiled, containing the same filler sentences plus the original high- and low-stress versions of the twenty experimental sentences, balanced across tapes for stress level.

The sentences were presented to subjects over headphones, and they were asked to press a button as soon as they heard a word beginning with the sound specified as target for that sentence. At the end of the experiment they were given a comprehension test; the scores were high, indicating that the subjects were indeed understanding the sentences.

The results for both the spliced and the original, or unspliced, sentences are presented in Table I.

Table I

Mean RT (msec) to presence of target phoneme

	Predicted high stress	Predicted low stress
Spliced sentences	378	417
	High Stress	Low Stress
Unspliced sentences	294	403

Two separate analyses were performed on the results. A mean RT score for each subject for each condition was computed and an analysis of variance performed on these scores; and a mean RT score for each sentence for each condition was computed and a similar analysis carried out. The combined results of these analyses allowed computation of the min F' statistic (Clark 1973).

As can be seen from Table I, the predicted-high-stress targets elicited faster RTs than the predicted-low-stress targets. This difference was significant, min F' (1,30) = 6.54, p<.025. The difference between the actual high- and low-stress target was in the same direction and nearly three times as large. Thus the effect of predicted stress level, while large and significant, does not account for the entire RT difference previously found between high- and low-stress words.

It did not appear that splicing of the tape had any effect in this study.[1] Although the sentences did not sound suprasegmentally normal, they also did not sound particularly strange. In order to see whether a predicted-high-stress word having actually lower stress sounded consistently stranger (or less strange) than a predicted-low-stress word having actually higher stress, the experimental sentences were played to a group of subjects (not the same ones as in the RT task) who judged their "oddness" on a 5-point scale. There was no significant difference between their ratings of the oddness of predicted-low-stress and predicted-high-stress sentences. In fact, there was also very little difference between the ratings given these sentences and the ratings given some normal sentences included in this judgment task as filler items, whereas some "weird" sentences also included, produced by splicing together into sentences single,

words or short phrases from many different contexts, received "oddness" ratings which were markedly higher.

In conclusion, this experiment provides solid evidence for the prediction of upcoming stress locations as an integral part of sentence understanding. An acoustically identical word was perceived faster when embedded in a suprasegmental context which predicted that it would bear high stress than in a context which predicted that it would bear low stress. Obviously, this difference cannot be ascribed to superior intelligibility of stressed words. Instead, the contextual variations must be the source of the difference. Exactly the same sequence of words preceded the target item in each case; the only difference lay in the intonation contour. We are therefore forced to assume that the sentence processing mechanism made use of this intonation contour to determine the relative stress level of items in the string before they actually occurred; further, special attention was apparently directed towards the high-stress items.

Presumably this focusing of attention is not without purpose; it must be assumed to facilitate the process of sentence understanding. What are the properties of stressed words which might aid the comprehension process, so much so that the processor finds it strategically useful to focus attention on them? On the one hand, as we have noted, these words have acoustic properties that make them more intelligible than other words in the string; on the other, they tend to correspond to points of high information in the sentence.

Either of these properties could be sufficient reason for it to be worthwhile to predict the location of high stress. Since stressed words are easier to decode, processing them faster might provide contextual aids in the decoding of neighboring words - and hence it would be reasonable to search actively for them, in order to speed the comprehension process as much as possible. Similarly, though, identification of the items bearing greatest information in the sentence might also serve to aid identification of neighboring items. Whether one or both of these possible reasons for the prediction of stress location are valid is a subject for further study.

Footnotes

[1] However it is planned to repeat the experiment using synthetic stimuli (which allow stress and intonation to be controlled independently of the segmental aspects of the sentence), thus eliminating any possible effects of tape-splicing.

References

Clark, H.H. (1973) The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

Cutler, A. and Foss, D.J. (1973) The Importance of Lexical Item Stress for Lexical Access. Paper presented to the 45th Annual Meeting, Midwestern Psychological Association, Chicago.

Foss, D.J. (1969) Decision Processes during Sentence Comprehension: Effects of Lexical Item Difficulty and Position upon Decision Times. Journal of Verbal Learning and Verbal Behavior, 8, 457-462.

Klatt, D.H. and Stevens, K.N. (1972) Sentence Recognition from Visual Examina-
tion of Spectrograms and Machine-Aided Lexical Searching. Proc. 1972
Conference of Speech Communication and Processing, Bedford, Mass.

Lea, W.A., Medress, M.F. and Skinner, T.E. (1973) Prosodic Aids to Speech Recog-
nition: III. Relationships between Stress and Phonemic Recognition
Results. Univac Report No. PXI0430, St. Paul, Minn.

Lieberman, P. (1963) Some Effects of Semantic and Grammatical Context on the
Production and Perception of Speech. Language & Speech, 6, 172-187.

Shields, J.L., McHugh, A. and 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.

Department of Psychology
The University of Texas at Austin
Austin, Texas 78712