PROSODIC AIDS TO SPEECH RECOGNITION: VI. TIMING CUES TO LINGUISTIC STRUCTURE AND IMPROVED COMPUTER PROGRAMS FOR PROSODIC ANALYSIS

Wayne A. Lea, et al

Sperry Univac

Prepared for:

Advanced Research Projects Agency

31 March 1975

SPERRY UNIVAC
DEFENSE SYSTEMS

PROSODIC AIDS TO SPEECH RECOGNITION:
VI. TIMING CUES TO LINGUISTIC STRUCTURE AND IMPROVED COMPUTER PROGRAMS FOR PROSODIC ANALYSIS

by

Wayne A. Lea
Dean R. Kloker

Defense Systems Division
St. Paul, Minnesota
(612) 456-2434

Semiannual Technical Report Submitted To:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Director, IPTO

31 March 1975

Report No. PX 11239

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the authors, and should not be interpreted as necessarily representing the official policies, either expressed, or implied, of the Advanced Research Projects Agency or the U.S. Government.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Univac Defense Systems Division P. O. Box 3525 St. Paul, Minnesota 55165	Unclassified 2b. GROUP

3. REPORT TITLE

Prosodic Aids to Speech Recognition: VI. Timing Cues to Linguistic Structure and Improved Computer Programs for Prosodic Analysis

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Semiannual Technical Report: 1 September 1974 – 28 February 1975

5. AUTHOR(S) *(First name, middle initial, last name)*

Wayne A. Lea
Dean R. Kloker

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
31 March 1975	70	23

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
DAHC 15-73-C-0310 b. PROJECT NO.	Univac Report No. PX 11239
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d	None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Advanced Research Projects Agency 9400 Wilson Boulevard Arlington, Virginia 22209

13. ABSTRACT

Computer programs for detecting syntactic boundaries (BOUND3) and locating stressed syllables (STRESS) have been supplied to ARPA contractors. Descriptions of the operation and performance of these programs are provided, and suggestions for further improvements in these programs are given.

Experiments were conducted on various timing cues that correlate with phonological and syntactic phrase boundaries. By defining a group of syllables whose vowels and sonorants are lengthened at least 20% above their median values, we were able to locate 91% of the phonological boundaries between phrases, that had been perceived by listeners who heard spectrally inverted speech. Also, time intervals between the onsets of stressed vowels were found to be longer when a phonological phrase boundary occurred between the vowels. These experiments demonstrated the ability to detect over 90% of the major phrase boundaries from either the interstress time intervals or the lengthening of vowels and sonorants. Another experiment demonstrated that the time interval between two stressed vowels inversely correlated with the percentage of phones that would be erroneously categorized by various automatic phonetic segmentation and labelling procedures. Interstress intervals, as a measure of speech rate, thus provide cues to which phonological rules (e.g., "fast speech" vs "slow speech" rules) may apply at various points in an utterance.

Further experiments are planned, dealing with carefully controlled studies of how prosodic information gives cues to the type of sentence spoken, the appropriate syntactic bracketing to assign, and the occurrence of subordination. Further studies of the BOUND3 and STRESS programs will be conducted, and improved versions of those prosodic analysis programs will be used to aid parsing and word matching in speech understanding systems.

DD FORM 1 NOV 68 1473

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Recognition						
Speech Analysis						
Linguistic Stress						
Prosodies						
Prosodic Feature Extraction						
Syntactic Boundary Detection						
Stressed Syllable Location						
Rhythm						
Disjuncture						
Durations						
Phonological Phrase Boundaries						

SPERRY UNIVAC
DEFENSE SYSTEMS

PROSODIC AIDS TO SPEECH RECOGNITION:
VI. TIMING CUES TO LINGUISTIC STRUCTURE AND IMPROVED COMPUTER PROGRAMS FOR PROSODIC ANALYSIS

by

Wayne A. Lea
Dean R. Kloker

Defense Systems Division
St. Paul, Minnesota
(612) 456-2434

Semiannual Technical Report Submitted To:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Director, IPTO

31 March 1975

Report No. PX 11239

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC 15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the authors, and should not be interpreted as necessarily representing the official policies, either expressed, or implied, of the Advanced Research Projects Agency or the U.S. Government.

PREFACE

This is the sixth in a series of reports on Prosodic Aids to Speech Recognition. The previous reports appeared as follows:

I.	Basic Algorithms and Stress Studies	1 October, 1972	PX 7940
II.	Syntactic Segmentation and Stressed Syllable Location	15 April, 1973	PX 10232
III.	Relationships Between Stress and Phonemic Recognition Results	21 September, 1973	PX 10430
IV.	A General Strategy for Prosodically-Guided Speech Understanding	29 March, 1974	PX 10791
V.	A Summary of Results to Date	31 October, 1974	PX 11087

This research was supported by the Advanced Research Projects Agency of the Department of Defense, under Contract No. DAHC15-73-C-0310, ARPA Order No. 2010. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

SUMMARY

Two computer programs for prosodic analysis have been delivered to ARPA contractors, and are being incorporated into speech understanding systems at Bolt Beranek and Newman, and System Development Corporation. One program ("BOUND3") is an improved procedure for detecting boundaries between major syntactic phrases from substantial fall-rise "valleys" in the contours of fundamental frequency versus time. This program has been improved from earlier Sperry Univac versions, by using more efficient procedures for finding valleys in the fundamental frequency contour, eliminating some false boundary detections by more strict requirements on the durations of falls or rises in fundamental frequency, and assigning confidence measures to each boundary detection.

The other program ("STRESS") represents a major milestone in Sperry Univac's efforts to provide prosodic aids to speech understanding. It is an implementation of a procedure for locating stressed syllables in continuous speech. This program includes procedures for finding the high-energy nucleus of each syllable in the speech and measuring the 'size' (energy and duration) of each nucleus. Those syllabic nuclei that are stressed are then found by a context-dependent analysis of energy and fundamental frequency contours. Within each major syntactic constituent delimited by the boundary detection program, a search is made for the earliest high energy chunk of speech (that is, a syllabic nucleus) during which fundamental frequency is increasing. Tests are made among all those syllabic nuclei that have rising fundamental frequency and that are near the peak fundamental frequency in the constituent. The first such nucleus whose energy (actually, sum-of-dB values in the nucleus) is greater than a threshold fraction (currently, about 62%) of the energy in subsequent nuclei is chosen as the stressed "HEAD" of the constituent. (If no nuclei have increasing fundamental frequency within them, the choice of the HEAD is based solely on relative energies or 'sizes' of the nuclei.)

After locating the stressed HEAD in a constituent, a straight "archetype line" is defined to approximate the gradual fall in fundamental frequency from the peak to the end of the constituent. Other stresses are assumed to be associated with high energy (long duration) nuclei near regions where fundamental frequency rises above the archetype line. Fundamental frequency must not fall rapidly within the stressed nucleus associated with the rise above the archetype line, and again a test of relative sizes of nearby candidate nuclei selects the stressed nucleus.

This program implements the procedures used in our previous hand analyses of stress patterns. However, a number of improvements and new tests are included in this computer implementation of the "archetype-contour algorithm". For one thing, two different measures of the size of syllabic nuclei are used, and allowance is made for cases when extreme values of energy alone or substantially rising fundamental frequency alone may cause a nucleus to be chosen as the stressed syllable. Also, when the archetype line covers a long time span and fundamental frequency drops substantially below the archetype line, an additional test allows stresses to be found on long-duration nuclei, even if the fundamental frequency doesn't rise above the archetype line. An additional test permits long-duration nuclei just before pauses in the speech to be found as stressed.

The STRESS program was tested with several speech texts for which we already had listener's perceptions of stress levels, plus results of applying two simpler stress location programs and the hand analysis with the archetype algorithm. On the average, 89% of the syllables perceived as stressed were found by the program, while about one out of five locations were 'false', in that they did not locate a syllable perceived as stressed. Over half of these false locations were found to be due to fairly prominent ('almost stressed') syllables, false boundary detections, and failures in syllabic segmentation. However, other errors were due to detailed inadequacies in the STRESS program, such as the wrong choice of candidate nuclei, problems with the archetype line, some long prepausal unstressed syllables which appeared stressed, and short nuclei and falling fundamental frequency contours that resulted from unvoiced obstruents surrounding short stressed vowels.

It is important to note that, while the STRESS program confuses about 15% of all syllables between the "stressed" and "unstressed" categories, listeners at their best performance confuse 5% of the syllables. Thus, while the program is open to some improvements, it is approaching the level of performance that listeners can attain. The program has also been shown to work considerably better than some simpler stress location programs.

Sperry Univac is currently cooperating with ARPA speech understanding system contractors in implementing the prosodic programs in their systems. Studies will soon be undertaken with subsets of the Sperry Univac speech data base, to determine

how the constituent boundaries detected from fundamental frequency contours are moved as the position of the first stress in the constituent is moved, and how the stress location program performs in locating the stresses at various positions within a sentence.

Another major effort has been concerned with experiments on timing cues to linguistic structure. In one experiment, five sentences per speaker were selected from the speech of six individuals who participated in simulations of computer interactions. The utterances were distorted by spectral inversion and presented to five listeners who marked stressed syllables, and the locations and types (normal or hesitation) of phonological phrase boundaries, using only the prosodic cues remaining in the signal. Vowel and sonorant durations (with and without aspiration) were measured from spectrograms, and then declared stressed or unstressed based on the perceptions. Exploring the hypothesis that large increases in phonetic duration are syntactically determined, perceived boundary locations were compared with preceding segments which were 20% above the median length for that segment type. Using a rule which groups lengthened syllables, and from the lengthened group predicts phrase boundaries, 91% of the perceived boundaries were predicted. Of all the perceived phrase boundaries, those before silences longer than 200 milliseconds were more reliably predicted by lengthening than boundaries not at long silences. Locations perceived to be normal phonological phrase boundaries were more reliably predicted than those perceived as hesitations. Of the predicted boundary locations not perceived by listeners, some mark major syntactic boundaries, but most are at minor syntactic breaks, notably between modifiers and nouns, and after prepositions. The results also suggest that speaker differences and style variations may be important.

In another experiment, the question was whether or not one could detect major phrase boundaries from timing of prosodic features alone (such as onsets of syllabic nuclei found from energy contours), without the need for a prior determination of the phonetic sequence or the detection of lengthening of phonetic segments. We found that, in read sentences and paragraphs, as well as simulated man-computer interactions, time intervals between onsets of stressed vowels ("disjunctures") clustered near mean values around 0.4 to 0.5 second, with standard deviations of about 0.2 second. Contrary to published hypotheses, durations of disjunctures tended to increase about linearly with the number of intervening unstressed syllables. Mean disjuncture durations doubled when spanning clause boundaries, and tripled when spanning sentence boundaries.

Mean pause durations, as measured by durations of unvoicing, tended to be equal to or twice the mean interstress interval, for clause and sentence boundaries, respectively. Syntactically-dictated pauses thus appear to be one- or two-unit interruptions of rhythm. Long disjunctures also accompanied 95% of the perceived boundaries between phonological phrases, and were found useful in determining which of several minimally-contrastive syntactic structures had been spoken.

In a third experiment, we investigated how various measures of the rate of speech correspond with changes in phonological structure that should be handled by "fast speech" phonological and acoustic phonetic rules. The duration of the interstress interval was found to inversely correlate with the percentage of phones that were erroneously categorized by various available methods for automatic phonetic categorization. Other measures of speech rate, such as the number of syllables per unit time, were not as closely correlated with phonetic error rates. The interstress interval thus appears to be useful in predicting phonological rules that might apply to an utterance.

These experiments show further ways in which the location of stressed syllables could play an important role in speech understanding, and they expand the ways in which prosodic information could be used to determine syntactic and phonological structure.

In subsequent studies with subsets of the Sperry Univac speech data base, we will be investigating whether prosodic structures can provide cues to sentence type, contrastive syntactic bracketing, and subordination of phrases. Further investigations also will be conducted on timing cues to linguistic structure. Such experiments are expected to provide a better understanding of the relationship between prosodic patterns and various linguistic structures, and to provide ideas for improving the prosodic analysis programs. The improved programs will be integrated into speech understanding systems and tested for their effectiveness in aiding word matching and syntactic parsing procedures.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

1. INTRODUCTION

This is a report on work currently in progress in the Univac Speech Communications Group, under contract with the Advanced Research Projects Agency (ARPA). As a part of ARPA's total program in research on speech understanding systems, the research reported herein is concerned with extracting reliable prosodic and distinctive features information from the acoustic waveform of connected speech (sentences and discourses). Studies are being concentrated on problems of detecting stressed syllables and syntactic boundaries, doing distinctive features analysis within stressed syllables, and using prosodic features to guide syntactic parsing and semantic analysis.

Under previous contracts, Sperry Univac developed basic tools for prosodic and distinctive-features analysis, and conducted initial experiments dealing with prosodic patterns in connected speech. A review of that previous work was presented in a recent report (Lea, 1974d). In section 2 of this report, new and improved versions of the prosodic tools are described. These tools are currently being integrated into the BBN and SDC speech understanding systems. In section 3, some experiments on timing cues to linguistic structure are described. These experiments show some further ways in which prosodic information may be useful in speech understanding systems.

Section 4 provides a summary and plans for further studies. References are listed in section 5. An Appendix provides an explanation of the eighth-tone scale of fundamental frequency measurement that is used in the Sperry Univac prosodic analysis programs.

2. COMPUTER PROGRAMS FOR PROSODIC ANALYSIS

2.1 BOUND3 for Detecting Boundaries between Major Syntactic Constituents

Lea's research (1971, 1972, 1973b) showed that a decrease (of about 7% or more) in fundamental frequency (F_0) usually occurred at the end of each major syntactic constituent, and an increase (of about 7% or more) in F_0 occurred near the beginning of the following constituent. A computer program, based on the regular occurrence of F_0 valleys at constituent boundaries, was implemented as a FORTRAN program on the Sperry Univac speech research facility and tested with six talkers reading the Rainbow Script, two talkers reading a paragraph composed of only monosyllabic words, and a collection of 31 ARPA sentences involving eight talkers. The boundary detection algorithm was found to correctly detect 79% of all linguistically predicted boundaries in the readings of the Rainbow Script, 83% of all predicted boundaries in the "Monosyllabic" Script, and over 74% of all predicted boundaries in the 31 ARPA sentences. Almost half of the predicted boundaries that were not located were between noun phrases and following verbals; when these were neglected, about 90% of all other boundaries were correctly detected.

Some improvements in the boundary detection program were recommended in Sperry Univac Report PX 10232 (Lea, Medress, and Skinner, 1973a). These, and a few other refinements, have been incorporated into the BOUND 3 program that was distributed to ARPA contractors during December, 1974.

2.1.1 Use of Eighth Tone Scale

One improvement has been the use of an eighth-tone scale (see the Appendix) for representing F_0, rather than a simple Hertz value. This eighth-tone scale has been used in our previous plots of F_0 versus time (as in Figure 3 on page 9, Lea, Medress, and Skinner, 1973), but was not directly used in the earlier calculations of boundaries.

An important advantage of the eighth tone scale is that it shows the same change in number of tones for the same percentage change in F_0, so that the logarithmic scale of perceptual just-noticeable-differences is captured, and equivalent percentage changes in a woman's fundamental frequency (from 240 to 250 Hz, for example) and a man's fundamental frequency (from 120 to 125 Hz, for example) yield the same change in the number of tones (from 86 to 89 eighth tones, and from 38 to 41 eighth tones, respectively).

With the eighth tone scale, the 7% F_0 differences sought for as cues to syntactic boundaries can be found as simple differences of 5 eighth tones. Consequently, BOUND3 searches for simple decreases of 5 eighth tones followed by increases of 5 eighth tones. This simplifies the arithmetic, eliminating multiplications and divisions that were previously used.

2.1.2 Eliminating Some False Phonetically-Produced Boundaries

Another improvement is that each substantial rise or fall of F_0 must last for two or more segments, for the fall to qualify as the fall into a constituent boundary, or the rise to qualify as a rise after a constituent boundary. This helps eliminate false detections of syntactic boundaries due to a single point which is out of line. In particular, the high F_0 that occurs right after unvoiced obstruents, and the sudden dip in F_0 accompanying voiced obstruents (cf. Lea, 1973b), are less likely to trigger false detection of boundaries. Related to the demand for at least two-point duration in each new F_0 maximum or minimum is the setting of the thresholds for subsequent sufficient fall and rise. If, after finding a new two-point minimum, the threshold for sufficient rise after it is set at five eighth tones above the lowest attained value, then occasionally a single-point dip will trigger a false boundary detection. To eliminate this, the BOUND3 program includes a determination of the next-to-lowest value and requires the expected 5 tone rise to be measured from that value, not the absolutely lowest local minimum.

2.1.3 Assigning Confidence Measures to Boundary Detections

Confidence measures are another improvement suggested in Univac Report PX 10232 (pp. 24 and 25). They have been incorporated into BOUND3, and represent the most valuable improvement over previous versions of the program. The method of assigning confidences is based on those F_0 contour features that appeared relevant to being sure that an apparent boundary is in fact syntactically-produced (as contrasted with being phonetically-produced, as are the F_0 dips and jumps near obstruents). Regular boundaries between constituents within a clause are assigned a basic confidence measure of 5 units plus three times the magnitude of the subsequent F_0 rise. Thus, a 10-eighth-tone rise will get a basic confidence of $5+3(10)=35$. Then, if the time during which F_0 is rising is more than 3 time segments (30 ms), the basic confidence measure is incremented by another 5 units. If the next maximum F_0 was due to a high F_0 value

after unvoicing (that is, the maximum is preceded by unvoicing, and followed by an immediate F_0 fall), the confidence measure for the preceding boundary is reduced by the amount of drop in the next two segments plus the amount of drop in the next five segments. A rapid fall in F_0 after the maximum, followed by a leveling off, is thus penalized, since it is likely to be due to an unvoiced consonant, not a syntactic effect.

2.1.4 Detecting Sentence and Clause Boundaries

Clause boundaries, or embedded and unembedded sentence boundaries, are expected to be accompanied by long pauses (long stretches of unvoicing). Currently, a stretch of 35 unvoiced segments (350 ms) is required for a detection of a sentential pause. (This threshold could well be made a controllable program paramcter, to account for variable rates of speech and talker differences, but it currently is an absolute number.) Each pause is assigned a basic confidence measure (KONFIB) equal to the length of unvoicing. Then this KONFIB is incremented by any amount that the F_0 rise following the pausal boundary is greater than 20 eighth tones. If the lowest F_0 value near the pause is before the pause, any amount of F_0 rise beyond 20 eighth tones above that clause-final value is used to increment the confidence measure. This attention to F_0 changes at pausal boundaries is based on previous observations (Lea, 1972, pp. 74-5) that very large F_0 changes accompany clause and sentence boundaries. When such changes occur, we are very confident a syntactic boundary has occurred.

The CONFIB subroutine also assigns a "confidence in type" (KONTYP) measure, which says that when KONTYP is high, the associated boundary is very likely to be a boundary between sentences, or between clauses. The KONTYP measure is equal to the confidence in the boundary (KONFIB) at the pause, plus an increment equal to just how much over 30 eighth tones the F_0 rises after the boundary. (Thus KONFIB and KONTYP at pauses are equal unless at least 30 eighth tones rise occurs, in which case KONTYP is larger.)

The CONFIB subroutine also includes a provision for reducing the confidence of an F_0-detected boundary which is just before a pause. Often just before a pause, a brief terminal rise will occur, and an extra boundary will have been set at the valley just before the terminal rise. If an F_0 rise (without associated fall) occurs just before a pause, the confidence of that boundary just before the pause is reduced by 20.

It is worth noting that the chosen assignments of basic confidence measures assume that a 10 eighth-tone-rise in F_0 is about as good a cue to the presence of a syntactic boundary as is a 35cs stretch of unvoicing. From a study of the confidence measures obtained with the above-mentioned texts, a boundary with a confidence below 30 or so could reasonably be rejected, while one about 60 or more is very probably a true syntactic boundary. One with a confidence above 100 is most definitely a syntactic boundary, and probably is a sentence boundary.

2.1.5 Evaluating the Performance of the BOUND3 Program

The boundary detection program has been repeatedly tested in previous experiments (Lea, 1972, 1973a; Lea, Medress, and Skinner, 1973) and shown to correctly detect about 90% of all major syntactic boundaries predicted by an independent syntactic analysis. The improvements incorporated in the latest version are expected to slightly increase the percentage of correct detections and reduce the number of false alarms that would be found in further tests. However, these differences are not expected to be dramatic. The main advantages of the improvements are the improved efficiency of computation,the reduction of phonetically-produced false alarms, and (of most interest) the availability of confidence measures.

Ultimately, the real test of the boundary detection process also will involve being able to predict by rule exactly which constituent boundaries will be marked by F_0 valleys. Also, the expected location of the boundary needs to be investigated further. Previous tests suggest that the detected boundary (at the bottom of the F_0 valley) often does not coincide with the timing of the boundary between the last word of the preceding constituent and the first word of the following constituent. Tests with the designed speech texts (Lea, 1974b,c) will permit the development of rules which specify where boundaries should occur.

2.2 STRESS, for Locating Stressed Syllables

A program for locating stressed syllables may be useful in speech understanding systems for any of several purposes: 1) locating anchor points around which the most reliable automatic phonetic analysis and word matching can be accomplished; 2) locating areas where surface phonetic structures correspond most closely with underlying phonemic structures, thus increasing the reliability of word finding and word matching;

3) locating the most likely regions where important ("content") words are (since important words are usually stressed); 4) giving information pertinent to which phonological rules may apply (since stress plays a role in many phonological rules); 5) providing basic stress timing data that may be used to determine rhythm, disjuncture, and rate of speech; and (6) guiding syntactic and semantic hypotheses (since stress patterns relate closely to syntactic and semantic structures).

The program STRESS represents quite an improvement over the original hand-algorithm described in our previous reports (Lea, Medress, and Skinner, 1972; Lea, 1973a; Lea, Medress, and Skinner, 1973; Lea, 1974a). The basic ideas behind the STRESS program were outlined in Univac Report PX 10146 (Lea, 1973a). These are detailed more in the flow charts of Figure 1a and 1b.

2.2.1 Syllabification

STRESS calls subroutine CHUNK, which finds high-intensity regions bounded by substantial dips in energy. These "chunks" are presumed to correspond with syllabic nuclei; that is, they are the central parts of syllables, consisting of vowels and (sometimes) non-vowel sonorants. The energy dips are presumed to be due to pre-vocalic or post-vocalic consonants, and thus they reflect the general areas where boundaries between syllables are presumed to occur.

There are several ways in which this syllabification is not ideally accomplished by CHUNK. For one thing, although the energy contour is band-limited (60-3000 Hz), there are still some cases where unvoiced obstruents (specifically, strident fricatives) give the high-intensity chunks that we are trying to identify as "syllabic nuclei". Such unvoiced chunks are not currently distinguished from the voiced syllabic nuclei by the CHUNK routine. Under Univac funding, a routine called SYLLB was implemented, which uses CHUNK, along with an independent voicing test (based on a very-low-frequency (60-400 Hz) energy function) to do a more accurate job of syllabification. The SYLLB routine has not been incorporated into STRESS, but, in a total prosodically-guided speech understanding strategy, syllabification by SYLLB ultimately will be done before STRESS is called, under the PROSOD executive to be described in section 4.3. Because no voicing decision is currently involved in CHUNK, a number of (admittedly redundant and inefficient) tests for voicing were incorporated within subroutines HEADER and OTHERS. These 'voicing tests' are based on determining whether a value of F_0 was found, not on an independent voicing decision.

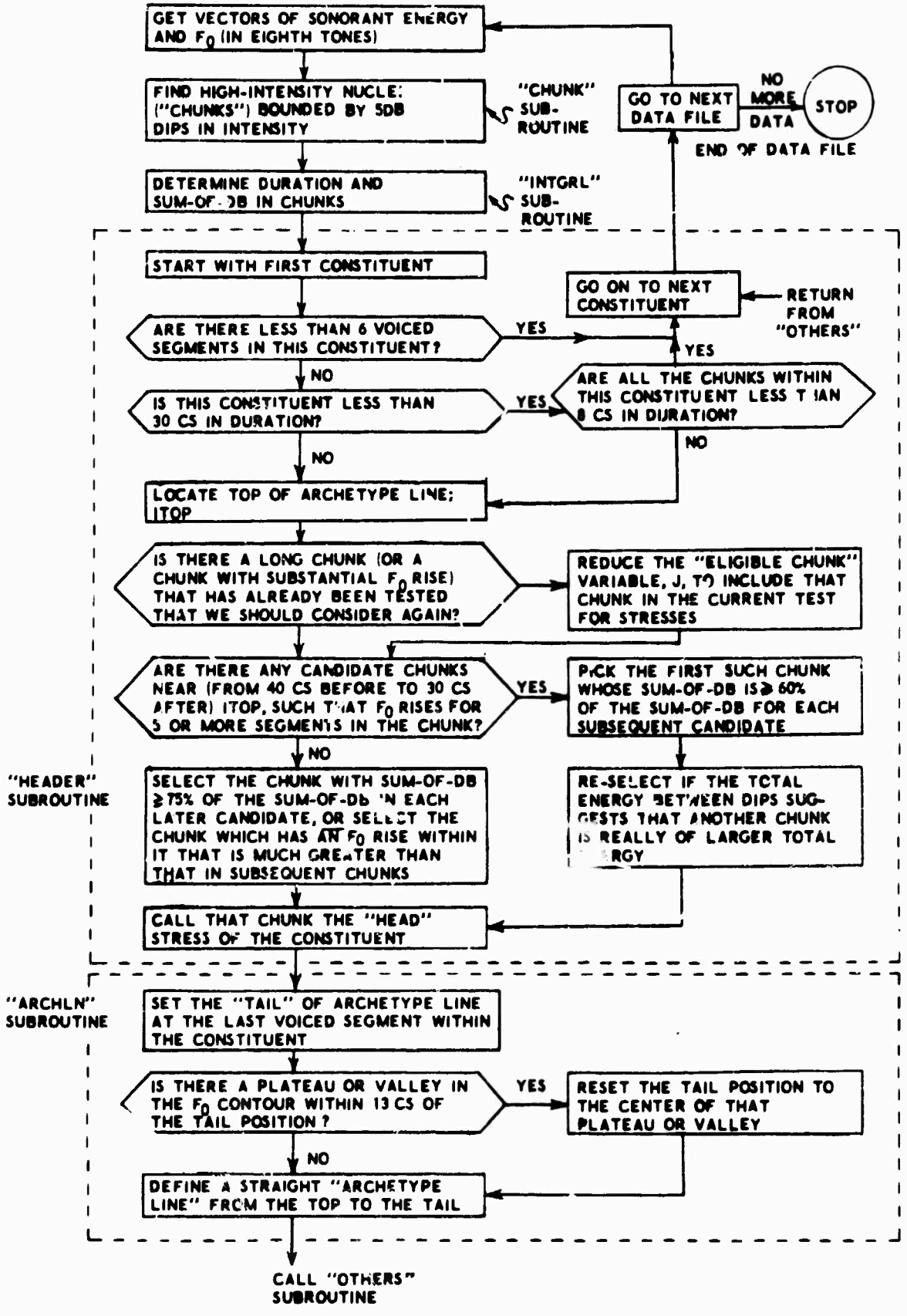


Figure 1a. Flowchart of the STRESS Program: Routines for Syllabification, Locating the Stressed HEAD in Each Constituent, and Defining the Archetype Line

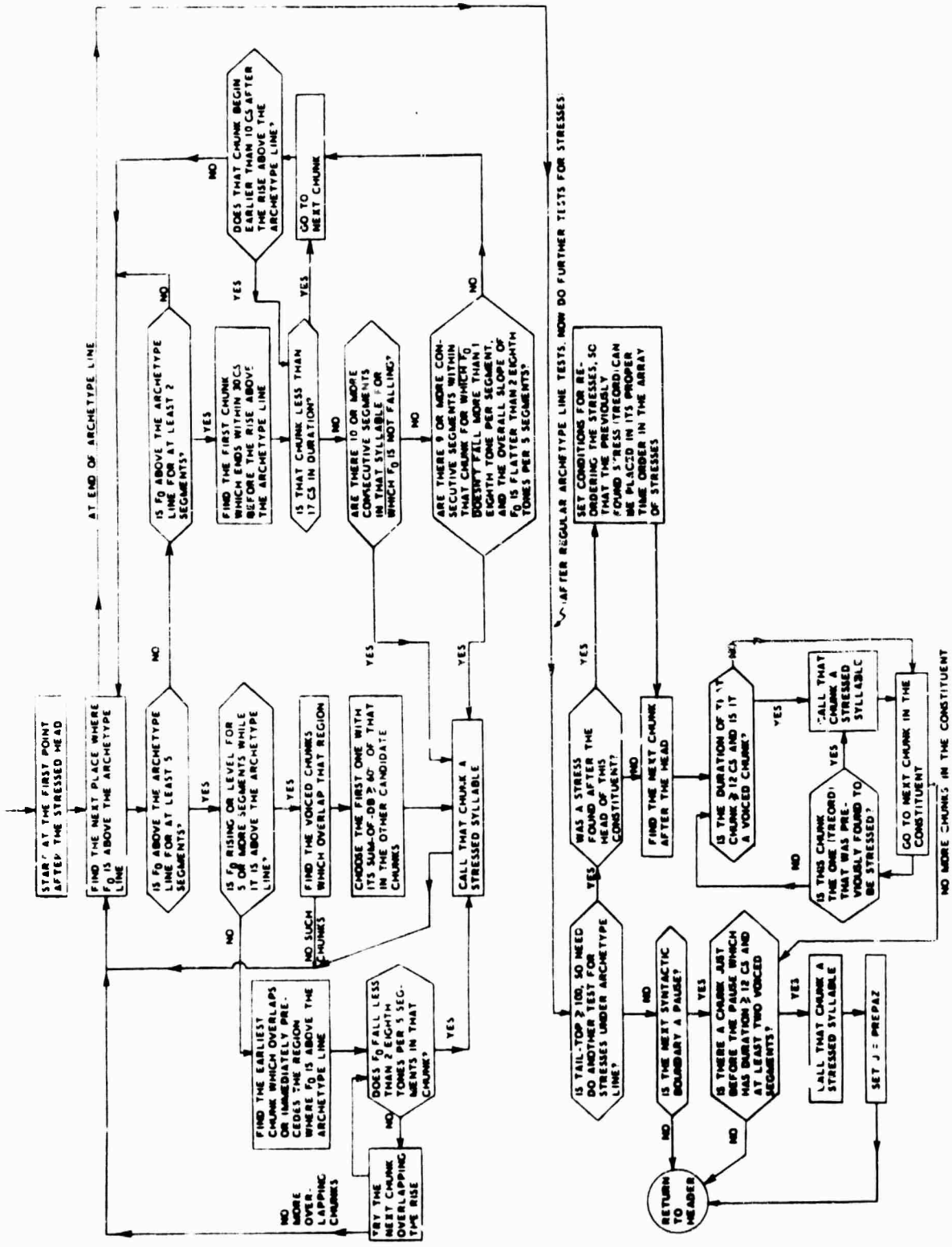


Figure 1b. Flowchart of the STRESS Program: The "OTHERS" Subroutine, for Locating All Other Stresses (other than the HEADS of Constituents)

While CHUNK is quite successful in finding most (over 90%) of all syllables in connected speech, there are some all-sonorant sequences which it fails to break up into syllables, due to the lack of a substantial dip in energy during the intervocalic sonorants. Thus, for example, in ARPA sentences LM3 and LS21, the sequence "the own-" or even "the owner" or "the owner of" may look like one long syllable. This 'inadequacy' in syllabification introduces false alarms in stress location, when sequences of unstressed syllables (like "do you", "-erence", "-cular", "for each", in the ARPA sentences) look like one long syllable. Also, the all-sonorant sequence "we were all" was found as one stressed syllable, even though it was perceived as having two stresses within it. This gave one "miss" in scoring stressed syllable locations.

CHUNK provides the 5 dB-down beginning NBEGIN(J) and end NEND(J) points of each chunk J, the position NPEAK(J) of the peak dB energy in the chunk, and the position NDIP (J) which is the bottom of the dip in energy between syllables.

Suggestions for refining the syllabification procedures will be given in section 4.2.

2.2.2 Measures of the 'Size' of the Syllabic Nucleus

Once syllabification is accomplished, STRESS than calls INTGRL to determine the duration of the syllabic nucleus and a measure of the "energy integral" for the nucleus. The duration LONG is simply the number of 10-ms time segments contained in the chunk, namely, LONG=NEND(J)-NBEGIN(J)+1. The "energy integral" measure is the sum of dB values within the chunk; namely, LARGE(J)=ISONOR(NBEGIN(J))+ISONOR (NBEGIN(J)+1)+ISONOR(NBEGIN(J)+2)+. . .+ ISONOR(NEND(J)). This measure is <u>not</u> a simple measure of the integral of speech power throughout the chunk. We tried using the actual energy integral, by converting the individual dB values for each of the time segments back to power, adding the powers, and converting back to dB. But this direct measure of energy integral worked very poorly in discriminating between stressed and unstressed syllables, for obvious reasons. For one thing, a syllable nucleus that is twice as long as another (with about the same segmental dB values) would only be 3dB higher in energy integral. A typical difference between a short unstressed syllable and a long stressed syllable might only be a change in energy integral from 60 to 61, 62, or 63 dB. An energy threshold for stress assignment is then difficult to assign. What is more, a 2 or 3 dB change in the intensity of a single point in the chunk can be almost as effective as a doubling of the duration of the chunk.

Instead, a measure is needed which is more sensitive to nucleus duration than to absolute intensity levels. Extensive previous studies have shown that duration is a better cue to stress than is intensity, and the Sperry Univac tests with a direct measure of energy integral vividly confirmed those results. The sum of dB values within a chunk (LARGE) was found to be an excellent combination of intensity and duration cues, in that it is heavily weighted toward duration cues, but doesn't entirely ignore intensities.

2.2.3 Locating the First Stress in Each Constituent

STRESS then calls subroutine HEADER to search for the first stress, or "HEAD", in each syntactic constituent. The beginning and ending of the constituent are provided by BOUND3, as is the position of maximum F_0 in the constituent.

Now, a couple of refinements on the original stress location algorithm come into play. HEADER first tests for whether there are at least 6 voiced segments (centiseconds, cs) in the constituent. If not, no test for stress is done in that constituent, since it is expected that any genuine constituent will have at least one voiced syllable which lasts more than 6 cs.

Any genuine "constituent" is also expected to be fairly long, containing several syllables or at least one long syllable. Consequently, HEADER tests for constituents that are shorter than 30 cs in duration. These may actually be short genuine constituents, or they may be "false" constituents resulting from local variations in Fo that are sufficient to trigger the syntactic boundary detection algorithm. HEADER thus requires a chunk of at least 8 cs duration (reflecting something like a minimum expected duration for a stressed chunk) within the constituent. If no such long chunk occurs, HEADER goes on to the next constituent, thus rejecting the idea that a stressed HEAD need be found in such short "constituents".

The first step in locating the HEAD in an acceptable constituent is to set some bounds on where to search. Certainly we look no farther back in time than the end of the last stress previously located. This is determined by LTEND, the time at the end of the last stressed chunk previously located. Some of the syllables right after that previous stress are also ruled out, or accepted as possible by adjustments of the chunk variable J. Then, HEADER will set some more stringent bounds on the earliest point where the stressed HEAD could be.

The most important definition of the search region for finding HEADs is to reject those chunks which begin more than 40 cs before the position of maximum F_0 in the constituent, or which end more than 30 cs after that maximum F_0 point. This asserts that somewhere near the maximum F_0 point in a constituent there is a stressed HEAD. It is most likely to be associated with the rise in F_0 from the previous constituent boundary up to the maximum F_0 point.

Thus, a number of chunks are initial candidates for possible stress assignment. HEADER than looks for those chunks which have 5 or more points where F_0 is rising, by determining how much F_0 has risen in the previous 5 time segments, for each time segment within each chunk. Each chunk which has a 5-point F_0 rise is considered an acceptable candidate for first choice as the stressed HEAD. The total amount of F_0 rise within a chunk is also calculated for each chunk.

Then the acceptable candidate chunks are compared, and that chunk is chosen which is the earliest one which has at least 62% of the "energy integral" LARGE of all subsequent acceptable candidate chunks. (The 62% is empirically derived; it had been set at 60% in the original algorithm, but slightly better results were obtained with the 62% threshold.) As of this point, a single choice for the HEAD has been selected; namely, stress is assigned to the first chunk with rising F_0 which is at least 62% of the energy of subsequent chunks with rising F_0.

However, it was necessary to allow a couple of other conditions to occasionally overrule the preliminary choice of the HEAD. If the total amount of F_0 rise within a candidate chunk is more than 10 eighth tones greater than that in an earlier chunk chosen in the preliminary choice, then that later chunk with very large F_0 rise is chosen as a more likely stress candidate. This revision of the choice of the stressed HEAD is based on the notion (Bolinger, 1958; Lea, 1973a) that large F_0 rises are associated with stress. This F_0 test was also empirically found to be necessary, to accomplish the correct selection of stressed HEADS, particularly in sentence-initial (post-pausal) constituents.

Another test which can overrule the selection of the HEAD chunk is a test of the "total energy" within a syllable. Sometimes, as shown for chunk J+1 in Figure 2a, a local increase of energy within a syllable can result in a short chunk riding on the top of a much longer high-energy region corresponding to the total syllabic nucleus. Another (earlier or later) chunk may be longer, and appear to have a larger sum of dB within its 5-dB down portions. The LARGE test might then select the wrong chunk for the stressed

HEAD. However, if a "total-energy" is properly defined for the syllable, one might be able to correct for this effect. This is done as follows. First, for each candidate chunk, a threshold amount of energy drop is defined as being two thirds of the distance down from the maximum energy in the chunk to the higher-energy of the two surrounding dips in energy. Thus, chunk J (actually, a temporarily chosen MAX chunk) is bounded by a previous energy dip at NDIP(J-1) and a following dip NDIP(J). Whichever dip has higher energy is used to define the threshold amount of energy drop that will be accepted within the syllable. Then, a threshold is set which is 1/3 of the distance up from the dip towards the maximum energy in the chunk. In Figure 2b, we see that for chunk J, NPEAK(J) has 60 dB and NDIP(J) has a low energy of 48 dB, so the threshold for the total energy measure is set at $48+(60-48)/3=52$ dB. The "total energy" measure LTOTAL is then defined as the sum of all dB values of sonorant energy in the syllable which are greater than or equal to the threshold value. A similar setting of thresholds and summing of dB is done in the other candidate chunks (such as chunk J+1 in Figure 2b). The chunk chosen as HEAD is then that one which has a total energy greater than 64% of all later acceptable chunks.

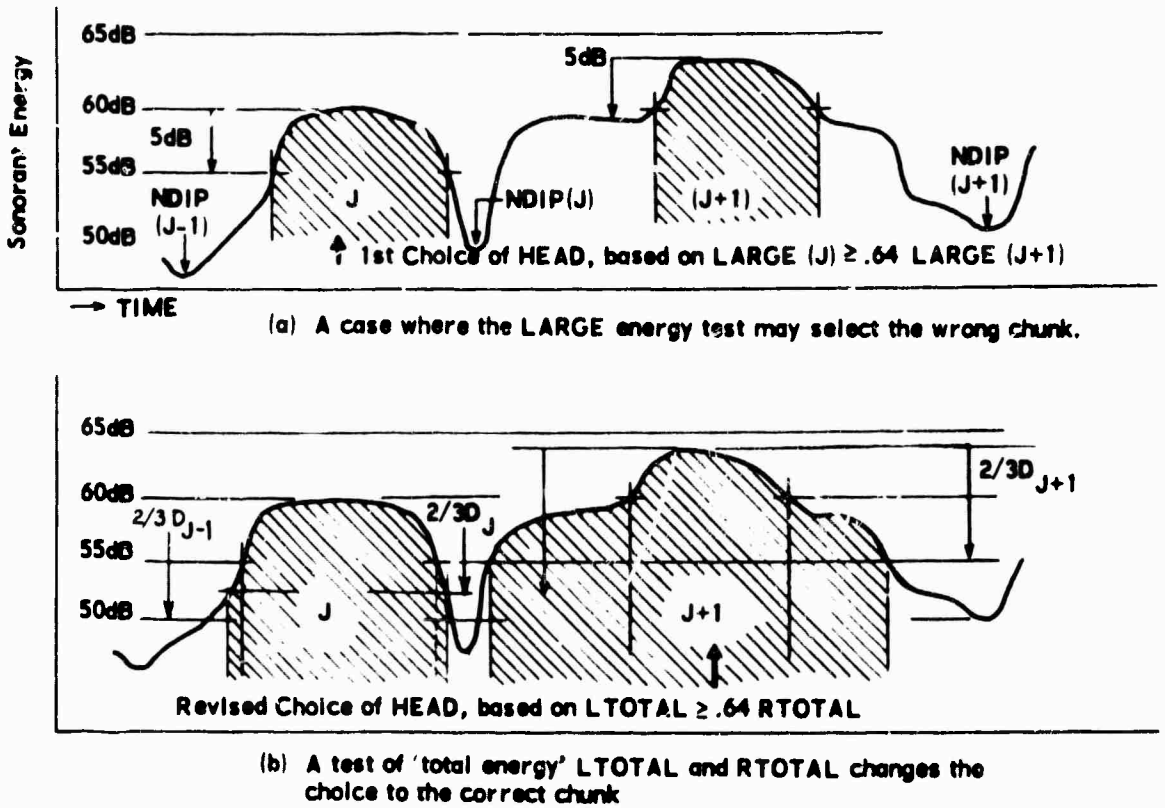


Figure 2. Revisions of the Selection of a HEAD Based on a 'Total-Energy' Test

This "total-energy" test doesn't override the test based on LARGE very often, and even then the earlier test with LARGE provides a single MAX chunk to compare all others with, thus making the "total-energy" test simpler.

These tests for the HEAD all work with those chunks which have a 5-point F_0 rise. There is another case which must be dealt with, however. Some constituents have no eligible chunks with such an F_0 rise, and a HEAD for the constituent must be chosen from among chunks with falling F_0 contours. (This may be due to the falling F_0 contours that follow unvoiced obstruents, even in stressed syllables. The fact that a syntactic boundary had been placed just before such syllables does suggest that F_0 was substantially lower at earlier times, and is now falling from higher values.) In this case, HEADER looks for the first chunk that is at least 75% of the energy of all subsequent chunks.

After the selection of the HEAD by one of the above-described procedures, it is checked for voicing (does it have at least two voiced points within it?) and, if it has not already been called a stress by a previous test (during the processing of the previous F_0-detected constituent), then it is placed into an array of stressed syllables, with the peak-energy point in the chunk specifying the "position" of the stressed syllable. The end of that chunk is then called LTEND, and the search for subsequent stresses will begin after LTEND.

It would seem very reasonable to also say that all tests of chunks for locating the next stress (within this constituent or the next) should then begin at the next chunk (J=IHEAD+1), right after the stress just located In the initial implementation of the program, such a demand was neglected and yet the results were quite good, even though the next test for a stress started with whatever J happened to be after all previous stress tests. (Remember, J, the testing chunk number, can be several chunks after the finally-chosen stressed chunk, since HEADER tests several chunks and chooses the earliest of the best candidates, and the test variable J could be equal to the number of the stressed chunk, or as much as several chunks later). Inserting an "improvement" to set J equal to the number of the located stress and then to increment it to try subsequent chunks, caused the loss of several stresses that had been previously located correctly and, instead, produced some false alarms (chunks then called stressed that weren't perceived as stressed). It appeared that one should generally eliminate from

consideration most of those chunks which have already been tried. However, there were some exceptions, where one did need to decrease J by 1, to test an earlier chunk. These exceptions required J-1 to be a long chunk, with energy (LARGE) quite a bit larger in J than in J+1, or with J having a large F_0 rise within it.

As we shall see in section 4.2, two of the major areas where one would hope improvements may be made in the stress program are in the initial guidelines as to which chunks to test (what regions the candidate chunks can be in) and how to best select the stressed syllable from among the candidate chunks. Another improvement (in efficiency, at least) might be to collapse the several tests for the stressed HEAD into one or two more direct tests (such as testing LARGE and total energy in one energy test, and combining energy and F_0 tests more directly).

2.2.4 Defining the Archetype Fo Contour

After locating the stressed HEAD in a constituent, HEADER than calls ARCHLN, to define the "archetype line" used in the search for other stresses in the constituent. Subroutine ARCHLN and function LFO provide a mathematical formula for a straight line on the eighth-tone F_0 plot, extending from the maximum F_0 point ("TOP") down to a point ("TAIL") near the end of the constituent. It is expected that if there are no stresses between the HEAD and the end of the constituent, this form of a steady drop in F_0 will be exhibited. However, if any streses occur between the HEAD and the end of the constituent, we would expect them to be indicated by local increases in F_0 above the falling archetype line.

To define the archetype line, ARCHLN first defines the TAIL, or bottom end, of the line. (The TOP has already been defined by HEADER.) The time of the TAIL is initially set at the position LIMIT of the next syntactic boundary, or at the last point of voicing LTV, if the next boundary is within a pause. Then, the TAIL is moved back to the middle of the first plateau or valley in F_0 that is encountered in moving back from LIMIT or LTV (not including the plateau or valley that the boundary was set in). The value of F_0 (in eighth tones) at the TAIL then determines the bottom of the archetype line, and the line extends between the point (TTOP, TONES (TTOP)) and (TTAIL, TONES (TTAIL)). The function LFO actually gives the value of F_0 at each point on the line defined by these end points.

This definition of the archetype line is not profound, but rather empirical. Previous results showed that, just before many boundaries, F_0 would drop quite fast, and the last few points would thus not be representative of the more gradual fall in F_0 occurring during earlier parts of the constituent. One way to capture the basic contour was to set the TAIL near the end of the gradual fall and before the final rapid fall in F_0. Usually a little plateau or valley could be found in that region, as one cue to the end of the general archetype shape.

In section 4.2, we will consider some refinements that might be introduced into the definition of the archetype line.

2.2.5 Locating Other Stressed Syllables in the Constituents

Subroutine OTHERS looks for other stressed syllables between the HEAD and the TAIL of the constituent. It is called by STRESS after the archetype line is defined. Its most basic test is to look for places where F_0 rises above the archetype line for 5 or more segments, and to find a nearby chunk with nonfalling (or at least not-very-rapidly-falling) F_0. OTHERS also includes a few other "last resort" tests for other stresses.

2.2.5.1 The Basic Tests for Other Stresses

OTHERS finds a place where F_0 is above the archetype line for 5 segments. When such a region is found, tests determine whether F_0 is rising for 5 or more segments (in which case, OTHERS initiates a search for a nearby candidate chunk for stress assignment) or not (in which case, OTHERS initiates a test for a nearby chunk in which F_0 is not falling too rapidly). If there are several candidate chunks with rising F_0 within them, and they are all in the vicinity of the rise of F_0 above the archetype line, OTHERS selects the earliest one whose energy (LARGE) is greater than 60% of all the subsequent candidates. If no chunks with rising F_0 occur near the rise above the archetype line, OTHERS selects that 'nearby' chunk in which F_0 doesn't drop more than two eighth tones per five segments within the chunk. The stress is then assigned, and then another similar test is conducted at the next place where F_0 rises above the archetype line.

The search region "in the vicinity" of the rise in Fo above the archetype line is defined rather narrowly. For a chunk to be the source of the F_0 rise above the archetype, it must either overlap with the region where F_0 is above the line, or it must be slightly earlier. The F_0 rise probably begin somewhat before it actually showed above the

archetype contour, thus reflecting a possibility of a slightly earlier position for the stress. This same form of overlap with the rise above the archetype line is applied for the "nearby" chunks to be tested for a two-eighth-tone drop per five segments, except that the immediately previous chunk also is allowed to be acceptable if it is fairly long (of duration greater than or equal to 12 cs).

As currently implemented, the test for a chunk with a slope more gradual than two eighth tones per five segments does not also include any energy test, to have all previous selections of chunks. It is not known whether such a test would help improve performance of the algorithm, but no cases were observed where such a test would obviously eliminate false locations or correctly locate stresses not currently being found. This may be worthy of further study.

This summarizes the basic tests for other stresses in the constituent. They are based on substantial rises above the archetype line accompanied by non-falling (or very slowly falling) F_0 within the stressed chunk. We now consider some other tests for stressed syllables, based on less stringent requirements. (Stresses found by these further tests may be ultimately assigned lower confidence measures, since they constitute more liberal conditions, allowing more chunks to be called stresses.)

2.2.5.2 Stresses Near Short Rises Above the Archetype Line

Sometimes the F_0 contour rises above the archetype line only briefly (for less than 5 segments), but a nearby stress is evident by a long-duration chunk with rising or slowly falling F_0 in its vicinity. If two or more points on the F_0 contour are above the archetype line, then OTHERS tests chunks from 30 cs before that rise above the archetype to 10 cs after, and determines if there are any chunks that are at least 17 cs in duration. If so, it finds any place within the total syllable (that is, between NDIP (NTEST-1) and NDIP(NTEST) for chunk NTEST) where F_0 doesn't fall from segment to segment, for at least 10 segments. This chunk (NTEST) will be called a stressed syllable. If, on the other hand, F_0 doesn't stay nonfalling for 10 segments within the syllable, OTHERS then asks if F_0 at least doesn't fall more than one eighth tone from one segment to the next, for 9 or more segments, where the segments must all be within the chunk. If so, and if the overall slope in those 9 segments if flatter than two eighth tones drop per five segments, OTHERS calls that a stressed chunk.

2.2.5.3 Stresses Under Long Archetype Lines

On rare occasions, the time interval between the TOP of the constituent and the TAIL will be so long that the straight line archetype will not capture all those places where F_0 rises due to stressed syllables. A special test has been included in OTHERS to look for additional stresses when the TAIL is at least 100 cs later than the TOP of a constituent, and no more than one stress was found by the other tests in HEADER. Any intervening chunks of duration 12 cs or greater will be accepted as stressed if they are voiced for six or more segments. (An additional test on F_0 slope might be appropriate here, but none has been incorporated yet.)

Since this test of long constituents may involve inserting a stress between two stresses already found (such as between the HEAD and a later stress found to have rising F_0 above the archetype), it demands special procedures for reordering the stresses, so they end up in consecutive order.

The test for intervening stresses in long archetype lines is particularly useful for monotonic speakers (such as the talker who spoke sentences designated LM3, LM13, LM24, etc., in the 31 ARPA sentences), since the boundary detector may fail to find significant F_0 fall-rise valleys corresponding to constituent boundaries in such speech. Long chunks can then still be found as stressed even though they show no rises above the archetype line.

2.2.5.4 Prepausal Stresses

Just before pauses, F_0 often drops very rapidly. The TAIL of the archetype line will usually be set within the last syllable of the constituent before the pause. Both effects make it very difficult to locate prepausal stressed syllables, since they don't have the nonfalling F_0 or the rises above the archetype line. Consequently, an additional test for stresses that may be in prepausal positions has been incorporated into OTHERS. This test merely says that the final chunk before a pause is stressed if it is at least 12 cs in duration and has at least two voiced points within it. This picks up some (but not all) prepausal stresses, but it does also give some false alarms, since prepausal unstressed syllables are longer than utterance-medial unstressed syllables, and they may appear to be stressed syllables by this test.

In summary, OTHERS has a first choice for stressed syllables: those with rising F_0 that overlap (or are near) the 5-point rises in F_0 above the archetype line. Next, it allows cases where the F_0 slope in the chunk is slightly negative. Then, it allows cases

where 17 cs-long chunks are associated with brief (2 segment) rises above the archetype line. For long constituents (TAIL-TOP $\geq$ 100), it finds other stresses as any long (12 cs) voiced chunks. Lastly, prepausal stresses are allowed, wherever a prepausal (partially-voiced) chunk is at least 12 cs long.

These tests seem somewhat redundant, and in section 4.2 we shall consider some ways to substantially improve the efficiency and performance of the program. Before doing that, we next consider the results in applying the program to various speech texts.

2.3 Evaluating the Performance of the STRESS Programs

Univac has previously obtained listeners' perceptions of stress levels and the results of hand analysis with the archetype contour algorithm, for several speech texts: the Rainbow Script, by talkers ASH and GWH; the Monosyllabic Script, by the same two talkers; and the 31 ARPA Sentences (Lea, 1973a; Lea, Medress, and Skinner, 1973). We also have results of applying two simpler stress location programs to those same texts (Lea, 1974a). It is appropriate, then, to ask how well the STRESS program does on those texts, and to compare it with the hand algorithm, the simple programs, and the listeners' perceptions.

2.3.1 Correct Locations by the STRESS Program

All these texts have been run through the STRESS program, and the results are summarized in columns C4 and F4 of Table I. The program was originally designed and tested with the Monosyllabic Scripts as design data, and the 94% overall performance with those texts is very good. In particular, since we have shown previously (Lea, 1973) that the human perceptions of stress show 5% confusions from time to time or from listener to listener, we cannot justifiably demand that the stress location algorithm obtain more than about 95% of all syllables perceived as stressed (or 97.5%, if confusions are equally split between missed stresses and false alarms). We are thus within a few percentage points of optimum expected performance for the correct location of syllables perceived as stressed in the Monosyllabic Scripts. (We shall see later that even the false alarm rate is low, when we consider other factors that the human listener can bring to bear that the STRESS program is unable to consider.)

When the STRESS program was applied to the Rainbow Script, the results were also quite good, with an average of 92% correct locations of perceived stresses.

TABLE I. RESULTS IN STRESSED SYLLABLE LOCATION

Speech Text	Number of Perceived Stresses (Majority Vote)	PERCENTAGES OF STRESSES CORRECTLY LOCATED				PERCENTAGES OF ALL LOCATIONS THAT WERE 'FALSE'			
		Chunk Duration Only	F_0 Rises Only	Hand Archetype Algorithm	'Stress' Program	Chunk Duration Only	F_0 Rises Only	Hand Archetype Algorithm	'Stress' Program
Monosyllabic Script:	(P)	(C1)	(C2)	(C3)	(C4)	(F1)	(F2)	(F3)	(F4)
Talker GWH	41	95%	83%	95%	95%	25%	23%	26%	19%
Talker ASH	41	95%	85%	90%	93%	24%	22%	18%	16%
Rainbow Script:									
Talker GWH	45	87%	78%	98%	93%	29%	20%	14%	11%
Talker ASH	51	73%	79%	84%	90%	20%	23%	17%	16%
31 ARPA Sentences	165	76%	73%	86%	85%	38%	26%	23%	26%

Performance was somewhat less with the 31 ARPA sentences. The reasons for the lower (85%) score in correct locations for the ARPA sentences include the more monotonic F_0 contours and the less distinct energy dips between syllables in that speech, but they also include the fact that listeners may hear stresses where they may be expected syntactically or semantically, even when they are not acoustically (or, at least, prosodically) prominent. A number of perceived stresses that were not located by the STRESS program involved short lax vowels surrounded by unvoiced consonants, yielding short chunks and falling F_0 contours. Also, a few perceived stresses were in syllables that the F_0 tracker failed to find voiced, such as utterance-final syllables with vocal fry or very low F_0.

2.3.2 False Locations by the STRESS Program

Column F4 of Table I shows the percentages of all locations by the stress program that were 'false'; that is, they pointed to chunks that were not perceived as stressed by the majority of the listeners. About one out of every five locations points to a syllable which the majority of listeners did not perceive as stressed. Table II provides a breakdown of the reasons for such 'false' locations, with the most common causes listed first.

An important point is that over one third of all 'false alarms' were due to syllables that were perceived as stressed by one listener (but unstressed by two other listeners). These borderline cases of almost-stressed syllables (line a in Table II) can hardly be characterized as 'false' locations in the same sense that some of the other ("totally-unstressed") locations are false. To a lesser extent, this is also true of the nine cases labelled "fairly prominent syllables", in line b of Table II. These included words like "lives", "these", "each", "me", and lesser-stressed syllables of compounds, like "-bow" of "rainbow", or "-word" of "keyword", or "-lite" of "troilite". Prepositions containing diphthongs (e.g., "by"; line h of Table II) also occasionally had the duration and other cues indicative of stressed syllables, even though they were not perceived as stressed. On the other hand, some stressed nuclei didn't appear stressed to the program because obstruents flanking the vowel make the nucleus short, and unvoiced obstruents before the vowel gave it a falling F_0 contour. In such cases (line j in Table II), a search of nearby chunks would result in the wrong syllable being chosen as the stress.

Another important reason for false alarms was the failure to separate all syllables in the speech, so that (as listed at line c in Table II) sequences of several unstressed or

TABLE II
SOURCES OF 'FALSE ALARMS' IN STRESSED SYLLABLE LOCATION

	MONOSYLLABIC SCRIPT	RAINBOW SCRIPT	31 ARPA
a. Syllables Perceived as Stressed by One (but Only One) Listener	9	5	13
b. Fairly Prominent Syllables	2	1	6
c. Failures in Syllabic Segmentation	2	1	5
d. False Constituent Boundary		1	5
e. Phrase – Final ("Tune II") F_0 Rise		2	3
f. Long Prepausal Unstressed Syllables			4
g. Wrong Choice of Candidate Chunk			
Due to Priority of F_0 Rise		2	4
Due to Total Energy Test		1	1
Due to Search Region Near Rise Above Archetype		1	1
h. Prepositions with Diphthongs	3		1
i. Problems with Archetype Line			4
j. F_0 Variations and Short Nuclei Due to Obstruents			3

reduced syllables (e.g., "so I", "bow is", "Do you", "-erence") would appear to be one long stressed nucleus. While there is not much that can be done to eliminate the various 'false' locations listed in the previous paragraph, this problem with syllabification errors is one that might be reduced or eliminated by a better approach to syllabic segmentation. Likewise, the false locations due to false syntactic boundaries (line d in Table II) might be eliminated by a refined boundary detection algorithm or by the use of the confidence measures associated with boundary detections provided by the BOUND3 program.

Two other phrase-structure-related sources of false alarms are listed in Table II, lines e and f. In several phrases marking incompleteness or special semantic or pragmatic structures (parenthetical "according to legend", high pitch on "-pha" of "alpha" in ARPA sentences CV1300 and CV2300, and "count where" in D7), a terminal rise in Fo occurred in the last syllable of the phrase, initiating a search for an extra stress. At prepausal positions, unstressed nuclei can be long enough to appear to be stressed, with the current (12 cs) threshold on durations of prepausal stress.

One of the more troublesome problems in stressed syllable location is how to choose the best candidate chunk when the initial cue to stress (namely, an Fo rise) is present. As shown in line g of Table II, the wrong chunk is sometimes chosen because it has an Fo rise within it, even when the actual stress may be in a nearby chunk (which could even be much longer and of much higher energy). This is due to the high priority assigned to Fo rises, so that once a chunk with rising Fo is found, that choice cannot currently be overridden by any energy or duration test on syllables that don't have the right form of Fo rise. The wrong chunk is also chosen a couple of times due to its having the highest total energy (but not necessarily the highest LARGE). Also, a couple of times the wrong chunk is chosen due to the manner in which the search region is defined near a rise above the archetype line.

Finally, there were four instances (line i of Table II) where false locations resulted from bad definitions of the TAIL or TOP of the archetype line, and the fact that, right after a stressed HEAD, Fo may be above the archetype line due not to another stress, but due to the HEAD. Since the program rules out the HEAD as a candidate chunk, it could then choose the next chunk as an extra stressed syllable.

2.3.3 Is the STRESS Program Adequate?

As noted before, listeners' perceptions of stress show about 5% confusions between stressed and unstressed categories. Adding the total number of stressed syllables that were not located by the STRESS program ("misses") to the number of occurrences of unstressed or reduced syllables being called stresses by the program ("false alarms"), and dividing by the total number of syllables in the speech texts, we find that the STRESS program confuses about 15% of all syllables between the "stressed" and "unstressed" categories. This is very encouraging performance, showing a fairly close approximation to the kind of consistency that can be obtained by listeners.

As we have seen, some of the remaining confusions involve "borderline" stresses, while others are due to specific weaknesses in the STRESS program. It must also be admitted that the listener brings to bear other information not available to the STRESS program; namely, segmental cues such as formant transitions and phonetic categorizations, plus knowledge of the phonological, syntactic, and semantic constraints of the language.

Also of importance is how well the STRESS program has captured the basic ideas intrinsic to the hand algorithm which it is supposed to embody. A comparison of columns C3 and C4 of Table I (page 18) shows that the STRESS program usually yields stress location scores at least as high as those found with the previous hand analysis (cf. Lea, 1973a; Lea, Medress, and Skinner, 1972). Perhaps even more important is the more nearly equal performance obtained with various speech texts. This makes one expect that new texts, perhaps even of different speech styles, may be handled with nearly the same level of performance. (However, it is clear from Table I that performance does drop some in going from written texts to the spontaneous speech of eight talkers involved in the ARPA sentences.)

It is also fitting to compare the performance of the STRESS program with the performance of other programs for stress locations. In Table I, columns C1, C2, F1, and F2, we find the comparable results in stressed syllable location by other (simpler) algorithms. On the whole, the STRESS program locates about 8% more stresses than a program ("INTGRL", Columns C1 and F1) based only on the selection of all chunks of duration 10 cs or more (cf. Lea, 1974a). The false alarm rate is also about 10% less for the STRESS program. This difference in performance is substantial, but it might not be enough to warrant the increased complexity of the STRESS program in all applications. It takes a lot of added tests to get that extra 15 to 20% improvement in overall performance.

Columns C2 and F2 of Table I show results in stressed syllable location by locating areas where Fo is 'rising' for fairly long time intervals (cf. Lea, 1973c). That program ("ONLYFO") generally worked poorer than the other two programs. However, Lea (1974) has previously shown that where the chunk duration program (INTGRL) failed, the Fo-rise program (ONLYFO) often succeeded, and vice versa. Consequently, combining the two cues, as is done in the STRESS program, helps improve performance and reduces sensitivity to the type of speech being handled.

Another important aspect of the evaluation of the STRESS program concerns its efficiency of operation. As noted in section 2.2, there are several somewhat redundant tests involved in HEADER and OTHERS that could be made much more efficient.

Finally, we may note in passing that the STRESS program is basically empirically derived. Extensive previous studies have shown that substantial Fo rises occur near the first stress in a major syntactic constituent, and that local rises above the archetype line usually accompany each other stress in the constituent. The theoretical basis for such observations has not been systematically developed. However, the extensive forthcoming studies with the large Univac speech data base (Lea, 1974b,c) may help provide a much more systematic understanding of such effects. Hopefully, a system of well-defined and experimentally-verified rules of English stress and intonation will be obtained. These may lead to a more theoretical, and hopefully a more efficient, formulation of stressed syllable location procedures.

2.4 Incorporating the Prosodic Programs Into ARPA Speech Understanding Systems

Listings and computer-readable forms (cards and digital tapes) of the BOUND3 and STRESS programs were provided to ARPA systems contractors, along with some sample results for a few speech data files. BBN made the programs available over the ARPANET. BBN has also integrated the boundary detection program with their version of the Sperry Univac F_0 tracking algorithm. BBN and Sperry Univac interacted on the evaluation of the boundary detection program's performance with ten BBN sentences. Issues of continued interest are exactly which constituents are demarcated by the program and where the detected boundary is located in comparison to the actual time between the two syntactic constituents. It appears from initial analysis that the detected boundary is usually located after the underlying syntactic boundary, at the position of the last obstruent before the first stressed vowel in the

following constituent. Further comparative studies will be done, to resolve such issues and to determine how the boundary detections might be used to aid the BBN parser.

SDC has also been working on incorporating a version of the boundary detection program into their speech understanding system. The SRI grammar includes specific places where prosodic information can be used to guide syntactic analysis.

Both BBN and SDC also expect to use the boundary detections as necessary inputs to the STRESS program. Sperry Univac has specific plans for interacting with BBN and SDC (over the ARPANET and by on-site visits), to integrate these prosodic programs into their systems. Early attention will be given to using stress locations to aid word matching, and to using boundaries and stress patterns to aid parsing procedures.

3. EXPERIMENTS ON TIMING CUES TO LINGUISTIC STRUCTURE

3.1 Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries

An experiment was conducted to examine the following hypothesis: vowel and sonorant lengthening is a cue to the phonological phrase structure in spontaneous English speech. First, speech was recorded from six people who interacted with a simulated computer, played by an unseen second person. Then, five listeners marked stressed syllables and phrase boundaries while listening to some of the interactions. Finally, these perceived boundary locations were compared with hand measured segment durations, and a rule was defined to predict the location of the perceived phrase boundaries from preceding lengthened segments.

The subjects interacting with the simulated computer issued instructions, questions, and commands. Their language was restricted only by the semantics of the task domain; and the resulting speech was spontaneous in style – including thoughtful pauses and hesitations. This speech style contrasts with that in previous studies of segmental lengthening (Oller, 1973; Lehiste, 1972; Klatt, 1975), which have used nonsense syllables, isolated phrases, or speech read from a text.

The second step was to obtain listener perceptions of stressed and unstressed syllables, and perceptions of the type and location of phonological phrase boundaries. For this study, five sentences were chosen from each of the six talker-computer interactions.

The phonological structure of the sentences is of particular interest in spontaneous speech, because the manner in which talkers group syllables and words into phonological phrases does not always correspond to the grammatical structure of their utterances. Similarly, syllables which should not be stressed according to the dictionary or to syntactic structure, may in fact be stressed by the talker. Thus, a listener asked to mark perceived stress and phonological boundaries may be biased by the location of the syntactic boundaries and the expected word stresses. For this reason, the 30 sentences were first distorted by spectral inversion, and then presented to the listeners. It has been shown (Blesser, 1972) that this kind of distortion maintains perceptions of acoustic cues to phonological structure – that is, to pitch, duration, and rhythm – while otherwise garbling the speech.

Figure 3 illustrates the format used in the listening test. The vertical bars and lines were already on the sheet given to the listeners; the slash marks and the letters P and H represent the perceptions of one person. Although they were told that the distorted speech was English sentences, listeners were unable to recognize more than a few of the words. Therefore, the markings must not have been influenced by semantics. Double bars on the sheet marked sententence boundaries, and single bars surrounded silences of at least 200 milliseconds duration. The relative silence locations were given to listeners, as a graphical aid in finding their place under the handicap of the distortion. Preliminary studies had shown that these silence gaps always signaled a boundary; and without some such graphical help, the perception task was too difficult for the longer sentences, which contained up to 30 words.

By means of a tape loop the listeners heard the sentences as many times as they wished. They marked all perceived syllables as stressed or unstressed, and all perceived boundaries as normal phonological phrases or hesitations. Hesitations were defined as interruptions of what would otherwise be a normal phonological phrase.

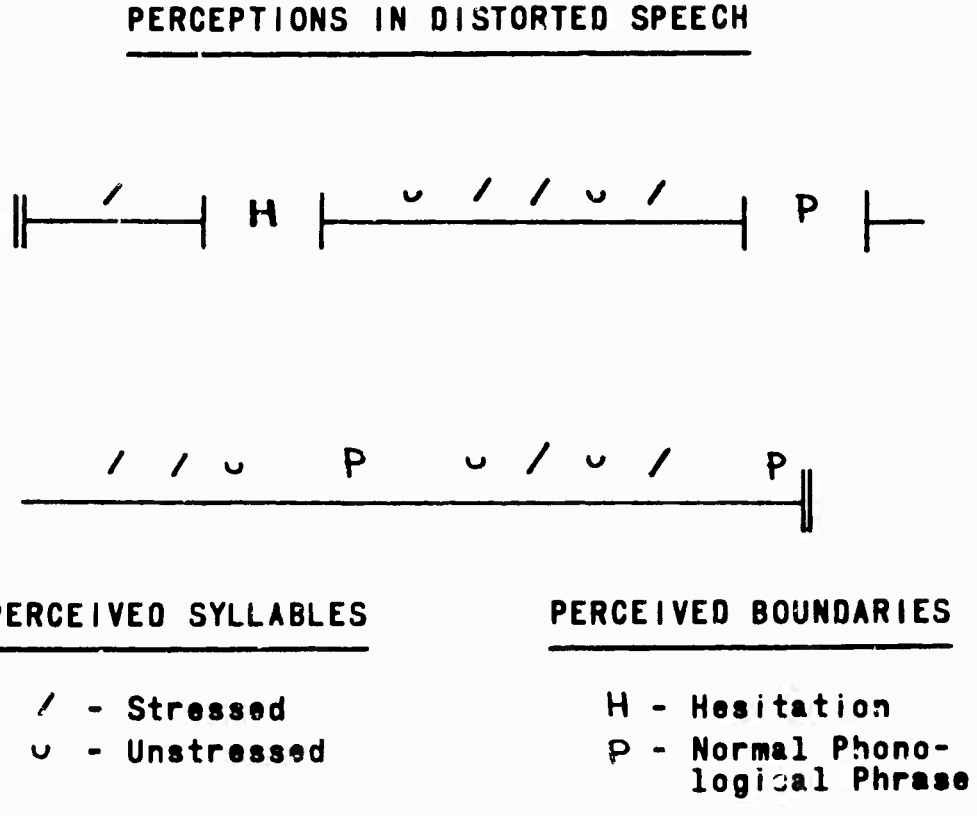
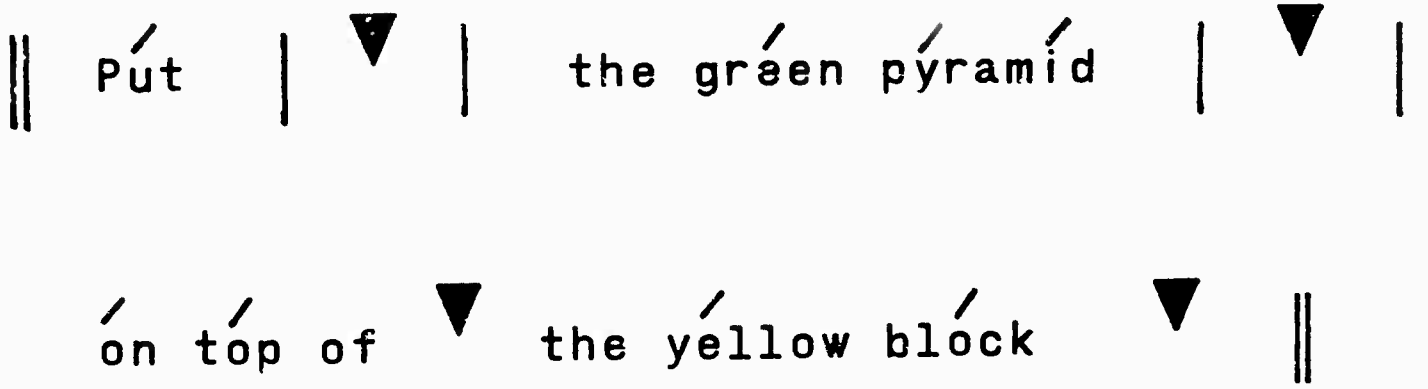


Figure 3. Listener's Perceptions of Syllables and Phonological Boundaries in Distorted Speech

As pointed out by Dennis Klatt (personal communication), any boundary perceptions dependent upon phonemic identification might be hampered by the garbled speech. That is, a listener may not be able to perceive a vowel as lengthened without the ability to identify the vowel. In addition, there may be cues to phonological structure based on the location of word boundaries; and word boundary perception might depend in part on phonemic identification. However, if no phonemic space is available to the listeners, then the notion of vowel and sonorant lengthening has no perceptual basis in the distorted speech. It seems more likely that listeners are able to normalize their perceptions to a phonemic space in the inverted spectrum, much as listeners readjust to the phonemic spaces of different talkers, especially after repeated listening to one talker. Blesser (1972) demonstrated that with practice in communicating via spectral inversion with another person, some people were eventually able to communicate with near normal perception of the speech.

Figure 4 illustrates the results of mapping listener perceptions in the distorted speech onto the actual sentence. The slashes mark stressed syllables, and the triangles mark perceived phonological phrase boundaries – ignoring for a moment the type of boundary. A location was counted as a phrase boundary when a majority of the five listeners marked it as such. Locations were identified by counting the number of perceived syllables and noting that reduced syllables were often missed. Figure 4 shows the location of a phonological phrase boundary after the word "of" instead of before the word, at the grammatical phrase boundary of the prepositional phrase. Such a misalignment is typical of this spontaneous style of speech. In the 30 sentences, about one-half of the perceived boundaries were at locations which are not major syntactic boundaries. Here, by a "major" boundary we mean one that is high in the surface structure tree of the sentence. This general statement is qualified later in Table IV, which shows the distribution by syntactic location for the perceived boundaries.

The locations of phonological phrase boundaries, such as in Figure 4, were expected to correlate with vowel and sonorant lengthening. Measurements of segment durations were made by hand from spectrograms, and the durations were tabulated separately by each talker. Following the lead of some recent work by Dennis Klatt, the median value of measured segment durations was chosen as the nominal, unlengthened value, and a percentage increase above the median was selected as the criterion for calling a segment lengthened.



/ Stressed Syllable

▼ Phrase Boundary

Figure 4. Listener Perceptions Mapped onto the Undistorted Sentence

Table III shows values of the median durations of the vowels and sonorants from one of the talkers. Separate medians were calculated for phonemes in stressed and unstressed environments. Segments were called stressed if they occurred in the first half of a syllable perceived to be stressed. Prevocalic and postvocalic sonorants were treated separately. In addition, segment durations were measured both with and without any preceding aspiration. The differences in results with respect to aspiration were small, but results were slightly better when aspiration was not included as part of the segment duration. The results in Table III do not include any preceding aspiration.

The median turns out to be a useful measure of the nominal, unlengthened duration of a segment, when at least half of the measurements are taken from non phrase-final positions. The underlying assumption, which holds true for this data, is that phrase-final lengthening is a much larger effect than other variations due to position or environment. Stress is the other major factor in length variation, and the reason why medians for stressed and unstressed segments were grouped separately.

As the criterion for calling a segment lengthened, a fixed percentage increase over the median works well, except for segments with a very short nominal duration. In these cases, the percentage gives such a small increase, that it may be swamped by other effects. Several percentages were tried. The best overall results were obtained when lengthened segments were defined as ones at least 20% above the median.

Previous studies using read speech and isolated phrases have generally implied that only the last syllable or two in a phrase are lengthened. Some production models of segment duration do predict lengthening more than two syllables back, but with less lengthening the further back you go. In this data, phrase-final lengthening occurs up to five or six syllables before the boundary. As a means of expressing this relationship between lengthened segments and phonological phrases, a group of lengthened syllables was defined by the following rule: group all adjacent lengthened syllables, that is, syllables with any segment lengthened at least 20% above the median for that segment type according to the maximum pattern of stressed and unstressed syllables:

u u / u / u u
(S S S S S S S),

where S is a syllable, and all but one of the syllables are optional.

Table III. Median Values of Vowel and Sonorant Durations

	Stressed	Unstressed
i	130	60
I	80	60
e	160	--
ɛ	110	(65)
æ	145	115
a	(185)	(60)
o	(120)	(70)
u	200	90
ʌ	90	--
ə	--	50
ɝ	80	95
m	--	(60)
-m	--	(70)
n	50	45
-n	--	60
l	45	(55)
-l	--	(70)
r	--	35
j	--	50
w	(50)	(60)

These results are for one of the six talkers.

The parentheses indicate only 2-4 occurrences of that segment.

The pattern shows the maximum length group. All but one of the syllables is optional, except that no more than two unstressed syllables in a row are allowed in the group formed. More explicitly, the following two patterns define the allowable groups; where parentheses indicate optionality:

```
  u u  /    u  /  u u     u u
((S)S) S ( (S) S (S(S))) ,S(S)
```

Thus, the word "the" with a lengthened vowel forms a group with only one unstressed, but lengthened syllable. The word "one" (with a lengthened nasal) forms a group as one stressed, lengthened syllable. The group "on top of" (with lengthened segments underlined) is formed from the pattern by deleting the first three and the last syllables. In the group "the four slot", the 1st, 4th, and last two syllables of the pattern are deleted. As it stands, the pattern could group as many as seven syllables, but in the 180 groups formed in the 30 sentences, the average number of syllables was only 1.5. Only two groups had as many as six syllables. These few longer groups are alternatively, and perhaps equivalently, analyzable as a slowing down in the rate of speech.

This rule for forming lengthened groups was developed partly because it fits the data well; but it is also motivated by the linguistic notion of a sense group, or tone group. In his book on intonation, Pike (1945) describes a phonological phrase group with a pre-contour, a single or double stress, and a post contour. The notion of a double stress group is exemplified in many noun-noun compounds, and adjective-noun pairs, such as "four slot", and "semicircle".

In Figure 5, the rule has been applied to the sentence shown earlier. The lengthened groups are indicated by the parentheses. A phrase boundary is predicted as the end of each lengthened group, with the exception that, if the end of a lengthened group is one unlengthened, unstressed syllable before a silence of at least 200 msec duration, then the predicted boundary is at the silence. A hit occurs where perceptions and predictions coincide. (In four cases, the perceived boundary location was only one unstressed syllable from the predicted boundary location, and because of the potential for error in translating listener marks to actual sentences, these cases were also counted as hits.) A miss occurs when a perceived boundary is not predicted, as shown after the word "put" in Figure 5. A false alarm occurs at a location predicted, but not perceived to be a boundary, as after the word "green". Most of the false alarms in the 30 sentences are at locations between a modifier and the following noun. (See Table IV.) Many of the perceived boundaries also are between modifiers and nouns.

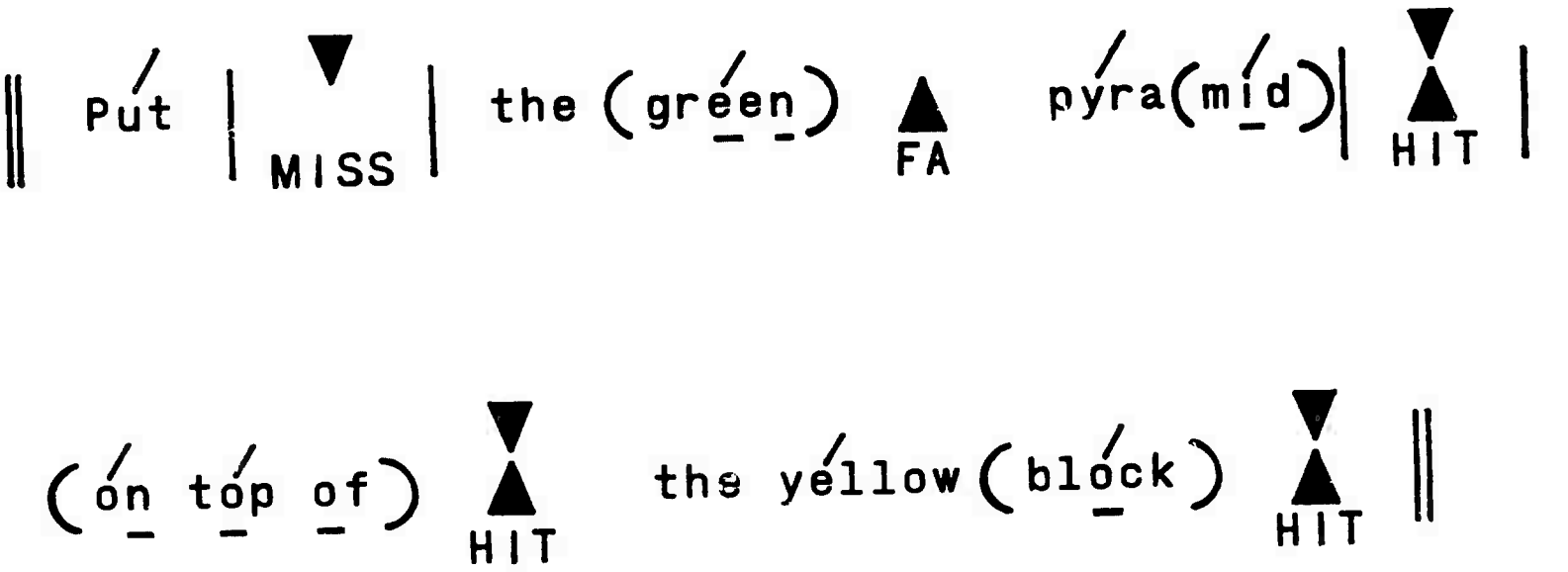


Figure 5. Predicted Phrase Boundaries from Lengthened Groups

Table IV. Comparison of Syntactic Locations

	Number of Occurrences	Location	Percent (No. of Occurrences) of: Perceived Boundaries		Predicted, but Not Perceived	
1.	46	V_NP	5%	(5)	12%	(5)
2.	21	NP_VP	5%	(5)	2%	(1)
3.	15	_[embedded clause]	12%	(13)	2%	(1)
4.	44	N_[non modifying phrase]	27%	(29)	5%	(2)
5.	38	N_[modifying phrase] * *PP or Reduced Relative	13%	(14)	12%	(5)
6.	185	[modifier] *_ (Adj.) N *Art., Adj., or N	20%	(22)	24%	(10)
7.	94	["first word"_] * *PP, embedded clause, or adverbial phrase	11%	(12)	21%	(9)
8.	48	"Other locations"	8%	(8)	5%	(2)
9.	--	"Non word-boundary"			17%	(7)
			100%	(108)	100%	(42)

(Results summed from six protocols, five sentences each.)

The second largest category of false alarm locations is "after the first word in a phrase"; mostly, in fact, after the preposition in a prepositional phrase. Again, many of the perceived boundaries also occurred in such a position, as in Figure 5, after the word "of".

As indicated in Table IV, seven such categories of locations are sufficient to describe 92% of the perceived boundaries and 78% of the false alarm locations. This suggests that many of the false alarms are at locations representing normal phonological phrase boundaries. Thus, the number of false alarms reported here is somewhat inflated with respect to any grammar that could account for the perceived phonological structure of this data.

Comparing the predictions with the perceptions for all 30 sentences, the results were quite encouraging. When only lengthened vowels were used to form lengthened groups, 81% of all perceived boundaries were predicted, while 31% of all predicted boundaries were false alarms. When vowels and/or sonorants are considered, the percentages of hits goes up to 91%, again with 31% false alarms. These results were obtained not counting sentence-final phrase boundaries, since we found that the average hit rate at the ends of sentences was only 63%, and other means were readily available for locating the ends of utterances.

We also calculated separate hit rates according to whether the perceived boundary was at a silence of at least 200 msec, or not. (Again, sentence-final locations were not included in the computations.) Phrases preceding long silences contained lengthened syllables 93% of the time, while phrases not before a long silence contained lengthened syllables only 80% of the time. It may be that boundaries marked with a long silence are stronger and more purposeful than others, and thus segmental lengthening is also stronger and more reliable.

Another factor to be considered is the type of the phonological phrase boundary. Recall that in the perception test using distorted speech, listeners marked boundaries as normal or as hesitations. Since the listening was done by five people, and the results were pooled, cases of disagreement were grouped into an "undecided" category. In the pooled results for the 30 sentences, 95% of all boundaries perceived as normal phrase boundaries were correctly located, while 89% of the "undecided" ones and 76% of the "hesitations" were correctly located. It is encouraging that boundaries marked as normal phonological phrases are more reliably predicted than those perceived as hesitations, that is, interruptions of normal phrases.

Thus, the data supports the hypothesis that vowel and sonorant lengthening is an acoustic cue to the phonological phrase structure, in spontaneous English speech.

3.2 Interstress Intervals as Cues to Phonological Phrase Boundaries

In previous reports (Lea, 1972; 1973a; Lea, Medress, and Skinner, 1973; 1975) arguments have been given for the development of a prosodically-guided speech understanding strategy, in which preliminary syntactic hypotheses are determined from acoustic prosodic data, without depending upon a prior phonemic analysis.

The experiments reported on in the preceding section suggest that boundaries between large phonological units can be determined from measuring the durations of phonetic segments. Yet, in keeping with the prosodically-guided philosophy, we may ask whether it would be possible to detect linguistic boundaries from timing information, without having to segment the speech into phonetic segments and then identify the lengthened segments with occurrences of boundaries. Are cues to linguistic boundaries also to be found from timing of simple prosodic information, such as the time intervals between syllables?

To answer this and other questions, a study was conducted of rhythm in the Sperry Univac speech texts (Lea, 1974a). Since English is said to be a stress-timed language, so that stressed syllables tend to occur at nearly equal intervals, and since stressed syllables have already been shown to play important roles in speech understanding, the time intervals between perceived stresses were analyzed. As discussed in an earlier report (Lea, 1974a, pp. 34-45), stressed syllables did tend, on the average, to be spaced about 4 tenths to 5 tenths of a second apart. Despite the general clustering of interstress intervals near 0.5 seconds, there was enough variations in interstress intervals (with standard deviations on the order of 0.2 seconds for each text; Lea, 1974a), so that the concept of English being stress timed clearly was not simply exhibited by nearly exact equality of all interstress intervals, regardless of other factors. In particular, the number of unstressed syllables between two successive stresses was shown to have a substantial effect on how they are spaced. Figure 6 shows a plot of the size of the interstress interval versus the number of unstressed syllables between the stresses, for the 31 ARPA sentences. A dot is shown for each occurring interval, at the coordinates of its interval size and number of intervening unstressed syllables. The average interstress interval increases almost linearly with each new unstressed syllable that is introduced between stresses.

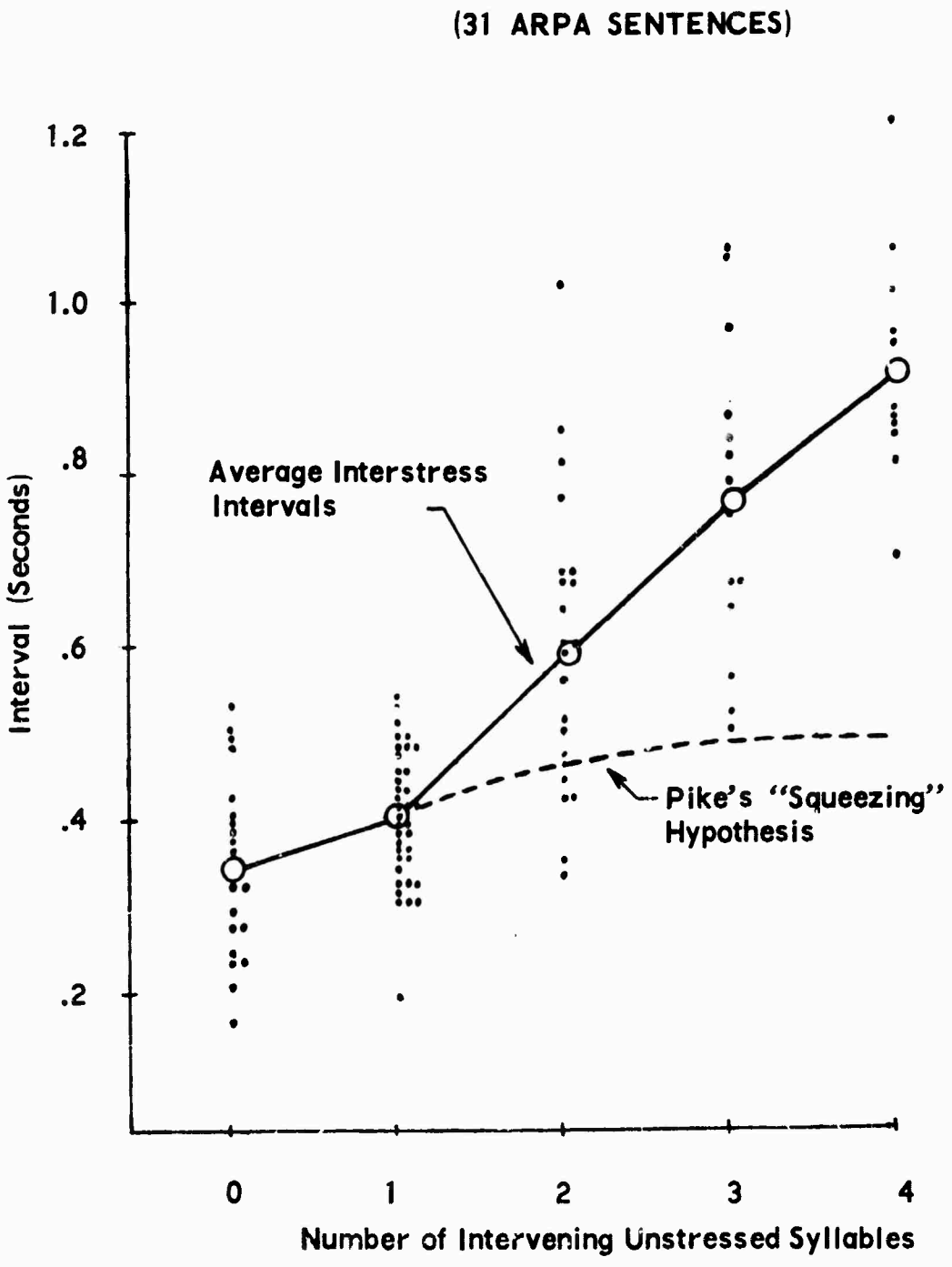


Figure 6. The Interstress Interval Increases with the Number of Intervening Unstressed Syllables

These results conflict with Kenneth Pike's (1945) hypothesis that equal time intervals between stresses may be a result of squeezing unstressed syllables closer and closer together as their number increases between two stresses. His hypothesis would predict (cf. Allen, 1968) that the interstress time interval would expand less and less as more and more unstressed syllables are added, as shown by the dotted line on the slide. For the 31 ARPA sentences, as shown in Figure 6, and for all the texts analyzed (cf. Lea, 1974a), no such tendency was found. Rather, each unstressed syllable (with the possible exception of the first one; cf. Allen, 1968; Lea 1974a) tends to add an equal increment to the interstress interval.

Figure 6 includes only those interstress intervals that did not span any pause in speech. As shown in the top left histogram of Figure 7, interstress intervals spanning pauses at clause boundaries tend to cluster at values near two times the average interstress interval found within uninterrupted speech. The dotted vertical lines show integral multiples of the 0.47 second mean interstress interval that was found for the uninterrupted intervals in the Rainbow Script (cf. Lea, 1974a, p. 35). The mean for intervals spanning clause boundaries is 0.99 seconds, which is quite close to twice the uninterrupted mean. Also, as shown at the top right in Figure 7, the duration of the pause itself clusters around a mean of 0.46 second, which is almost equal to the 0.47 second mean for uninterrupted interstress intervals. The pause at a clause boundary is thus a one-unit interruption of the speech.

Similarly, the interstress interval between sentences is shown at the bottom left of Figure 8. The mean interval is 1.43 seconds, which is very close to three times the interval of uninterrupted speech. This is due to a two-unit pause at sentence boundaries, as shown by the bottom right histogram in Figure 7. Again,the mean value is very close to an integral multiple of the interval in speech which is not interrupted by pauses.

Thus, syntactically-dictated pauses appear to be one- or two-unit interruptions of rhythm.

Durations of pauses provide only one of several ways in which syntactic boundaries relate to rhythmic structure. We also found that substantial syntactically-dictated valleys in fundamental frequency contours tend to occur at integral multiples of the average interstress interval (Lea, 1974, pp. 44-45). In addition, longer time intervals between stresses occur when a syntactic boundary occurs between the stresses, even if no pause occurs. As shown in Figure 8, the durations of interstress intervals tend to be small when only a word boundary intervenes, but durations are increasingly larger for boundaries between phrases, clauses and sentences.

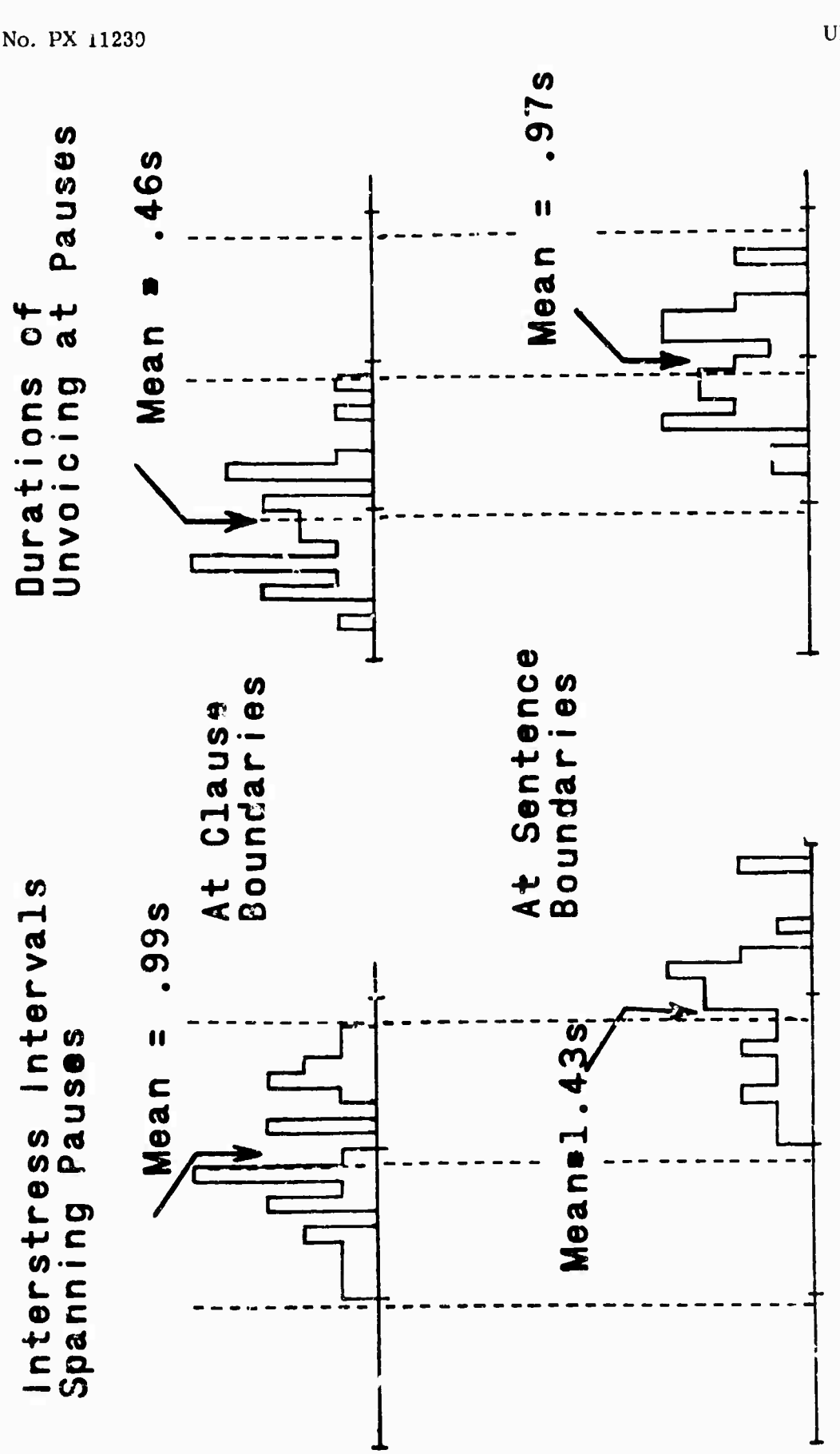


Figure 7. Syntactic Pauses Are Integral-Unit Interruptions of Rhythm

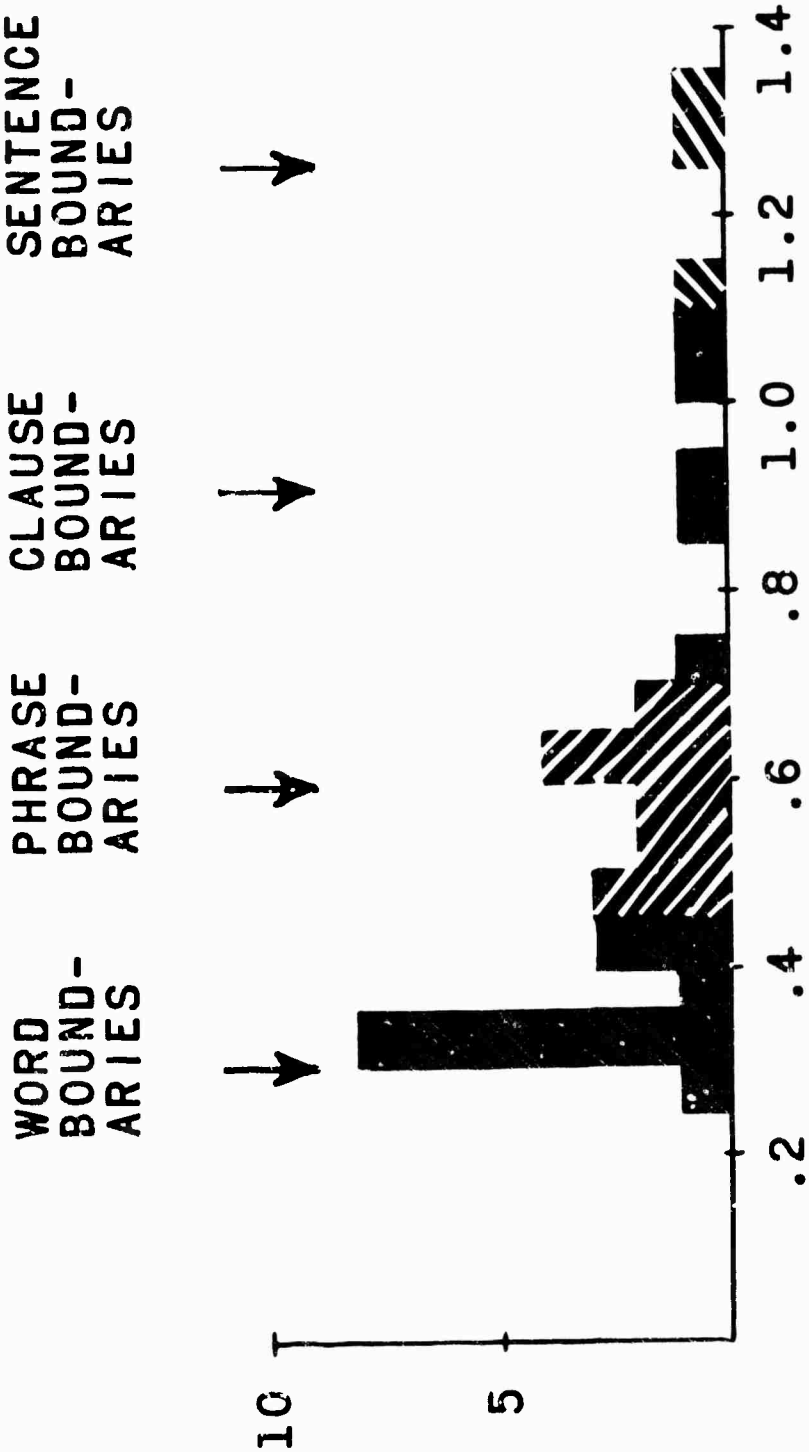


Figure 8. Longer Interstress Intervals Span Boundaries Between Larger Syntactic Units

Figure 9 further illustrates this point. Interstress intervals in the Monosyllabic Script, as read by talker ASH, are plotted along the abscissa, while the corresponding intervals (between the same syllables) for talker GWH are plotted along the ordinate. This gives a mark at the coordinates of the interstress intervals for the two talkers. Immediately obvious is how these corresponding intervals between the two talkers correlate very closely, so that the marks cluster close to a straight line. However, of particular importance here is not this consistency from talker to talker, but rather the effects of syntactic boundaries on how long the interstress intervals are. Interstress intervals were progressively larger as one progresses from minor syntactic boundaries, to major phrase boundaries, then clause boundaries, and then sentence boundaries. 'Minor' syntactic boundaries include those: between adjectives and nouns (shown by 'j' in Figure 9); between a noun and post-nominal modifier ("trek south"; symbolized by 'n'); between conjoined lexical items ('&'); between nouns and verbs ('v'); and between verbs and objects ('o'). 'Major' syntactic boundaries include those: between a verb and an adverbial phrase ('a'); between a noun and a relative clause (r); between a noun and a main verb, when there is an intervening auxiliary (m); and before prepositional phrases (p) and complements (t). Boundaries between full predicative clauses are shown by 'c', and those between separate sentences are shown by 's'. It would appear that long disjunctures (intervals between stresses) are a potential cue to major syntactic boundaries (as Lieberman observed, in 1967).

To further study the potential for automatically detecting syntactic boundaries from long disjunctures, a second set of experimental data was analyzed. From the experiment described in section 3.1 (cf. also Kloker, 1975), measurements were obtained of the time intervals between vowels in syllables which were perceived as stressed by five listeners. The listeners marked syllables they hears as stressed, and also marked where they heard phonological phrase boundaries. (Recall that these perceptions were obtained when listeners heard spectrally inverted speech, which preserves the prosodic information in the speech, but garbles the phonetic structure, so that the recognized words and English syntax cannot be directly applied to aid listeners in deciding where stresses and boundaries 'should' occur.)

Figure 10 shows the results for a typical sentence. Here the time interval between two stresses is plotted along the ordinate, and the time succession of stress-to-stress intervals is shown along the abscissa. Thus, the first interstress interval for sentence

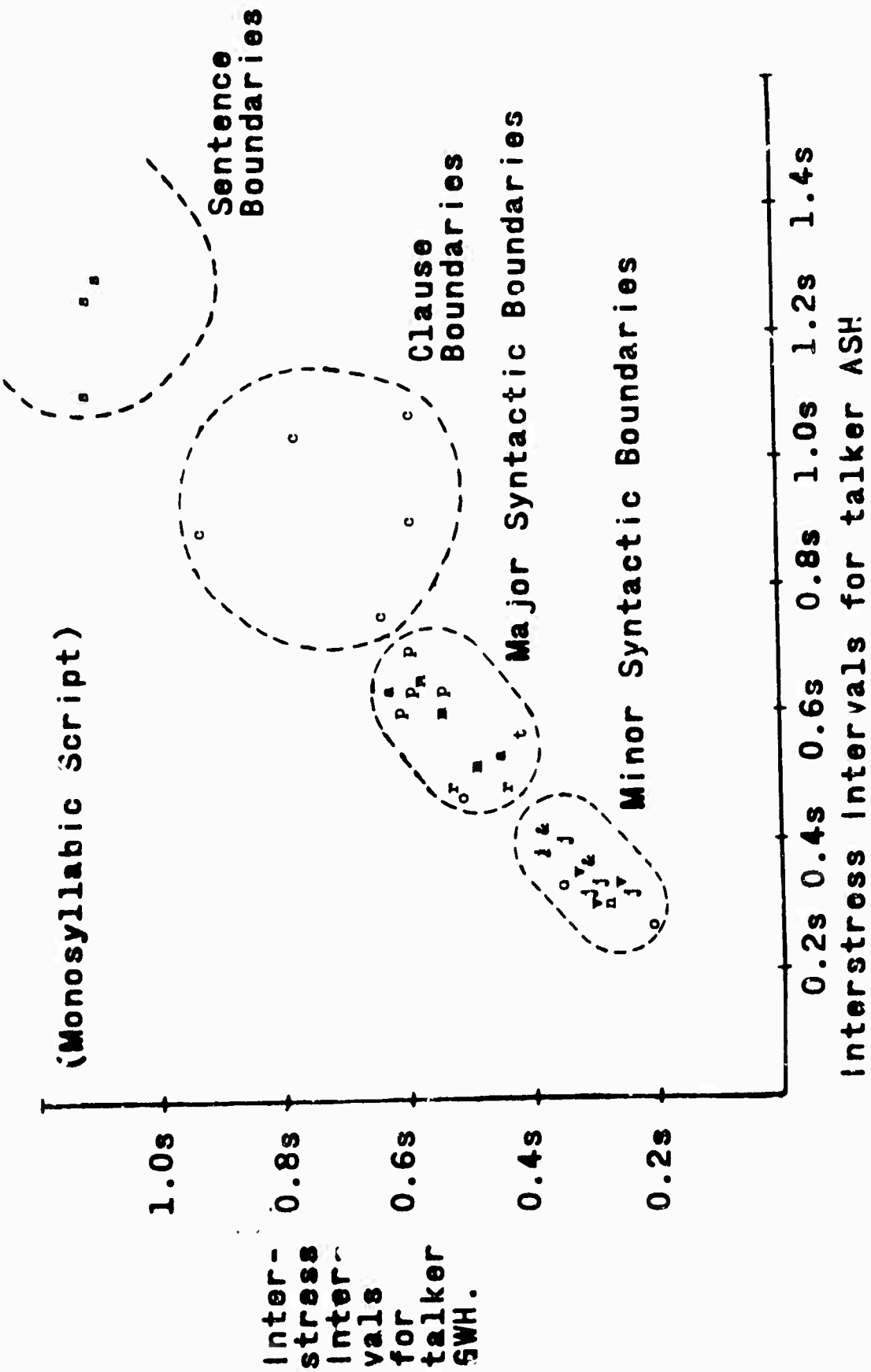


Figure 9. Interstress Intervals Are Consistent Cues to Syntactic Boundaries

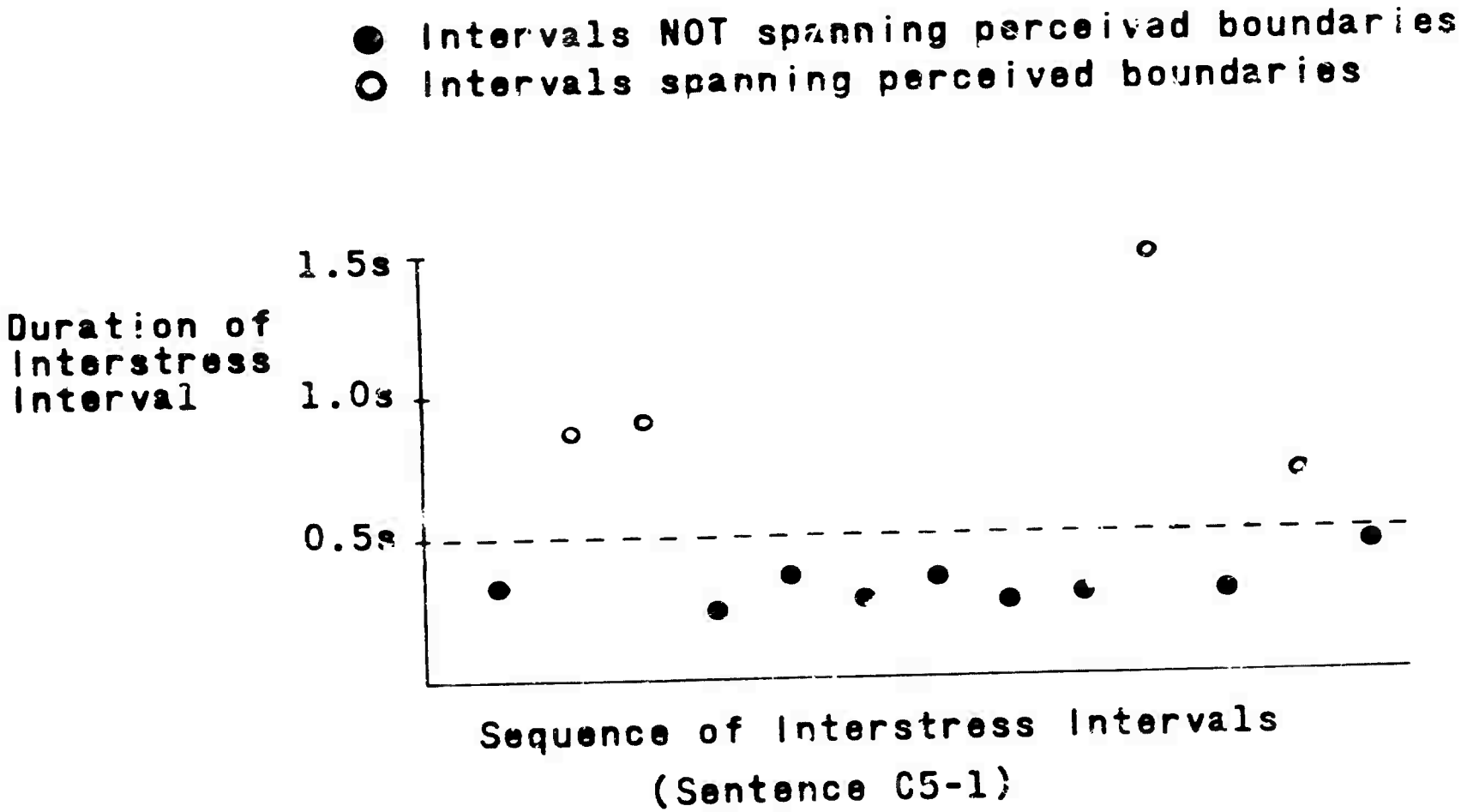


Figure 16. Long Interstress Intervals are Cues to Perceived Phonological Boundaries

C5-1 was 0.38 seconds in duration and did not span a perceived boundary, as shown by the leftmost black dot. The second interval was 0.90 seconds, and did span a perceived boundary, as shown by the leftmost open circle. The tendency for most intervals to be about 0.3 to 0.4 seconds in duration is vividly shown as one progresses through the series of intervals, shown by the black dots. This is a measure of the rate of speech. However, wherever there is a major phonological boundary, such that the listeners perceived a boundary in the spectrally distorted (inverted) speech, there usually is a very long interstress interval, as shown by the open circles in Figure 10.

These results suggested the following hypothesis: a phonological boundary is perceived wherever the interstress interval is greater than 5 tenths of a second. This hypothesis correctly detects over 95% of all the boundaries perceived (by a majority of the listeners). While it is true that 64% of these perceived boundaries were accompanied by silent pauses of at least 200 ms (so that a pause detector alone would find 64% of these boundaries), still it is encouraging that such a large percentage as 95% of the perceived boundaries can be detected from a simple threshold on the interstress duration.

Twenty three percent of all interstress intervals that were over 5 tenths of a second were not accompanied by perceptions of a boundary. However, these 'false alarms' in boundary detection did, in fact, span a major syntactic boundary in almost all cases. (Specifically, 25 of the 29 'false alarms' did involve major syntactic boundaries, even though the boundary was not perceived by listeners hearing the distorted speech.)

We may conclude that major phonological boundaries (which often correlate with underlying syntactic boundaries) can be reliably detected from long interstress intervals in the speech. This may be useful in a speech understanding system. Indeed, at Sperry Univac we analyzed several of BBN's structurally ambiguous sentences which had proved to be troublesome to the BBN speech understanding system. We found that prosodic cues, including pauses, time intervals between vowels, and stress patterns, could determine the correct syntactic structure, even when the parser and word-matching components of the system could not decide which structure was intended (Lea, 1974a).

3.3 Interstress Intervals as Cues to Applicable Phonological Rules

We noted in the previous section that the usual interstress interval that does not span a major syntactic boundary is one measure of speech rate In fact, we found that

the mean interstress interval correlated well with the total duration of a text, so that talkers that spoke a text (such as the Rainbow Script) more slowly than others had correspondingly longer mean interstress intervals. This is one indication that the interstress interval is a cue to speech rate. Now, it is well known that speech rate can have a significant effect on the phonological and acoustic phonetic structure of spoken sentences. Some phonological rules, called "fast speech rules", depend upon the rate of speech. Examples of such rules include vowel reduction and alveolar flapping rules. The question is: just what measure (or measures) of speech rate correlates well with the changes in phonological structure? Is it the time interval between stresses, the number of syllables per second, the number of phones per second, or perhaps the number of phrase boundaries per second?

The best correlation between a measure of speech rate and the variation in phonetic structures was found to occur for the interstress interval. This was established in a third experiment, as follows. Participants at the 1973 Speech Segmentation Workshop at Carnegie-Mellon University presented automatic segmentations of the 31 ARPA sentences into phonetically-labelled units. Their automatic labelling of segments was compared with a phonetic transcription provided by a linguist. Major discrepancies between the phonetic categories assigned by machine and linguist were considered to be errors. An investigation showed that most of these errors occurred where the interstress interval was short, as shown in Figure 11. The ordinate shows the percentage of all phones between two stresses that were erroneously categorized. Results are shown for each of the speech segmentation techniques used by research groups A, B, and C. The abscissa is the duration of the interval between the two stresses. Results as shown are actually the averages determined for interstress intervals quantized into 1 tenth of a second units. Thus, for group B, the average percentage of all phones which were erroneously categorized was 46% when the interval was between 0.2 and 0.3 seconds; the average error rate was 27% when the interval was between 0.3 and 0.4 seconds; etc., as shown by the points along the dashed line. For each group's method of machine labelling, the error rate was inversely related to the duration of the interstress interval.

This correlation between phonetic error rates and measures of speech rate was not as evident for the average (or, mean) interstress interval, or for other measures of speech rate, as shown by the correlation coefficients given in Table V The local measure of speech rate provided by the individual interstress interval usually yielded the highest-magnitude correlation coefficient, indicating that the phonetic error rate is

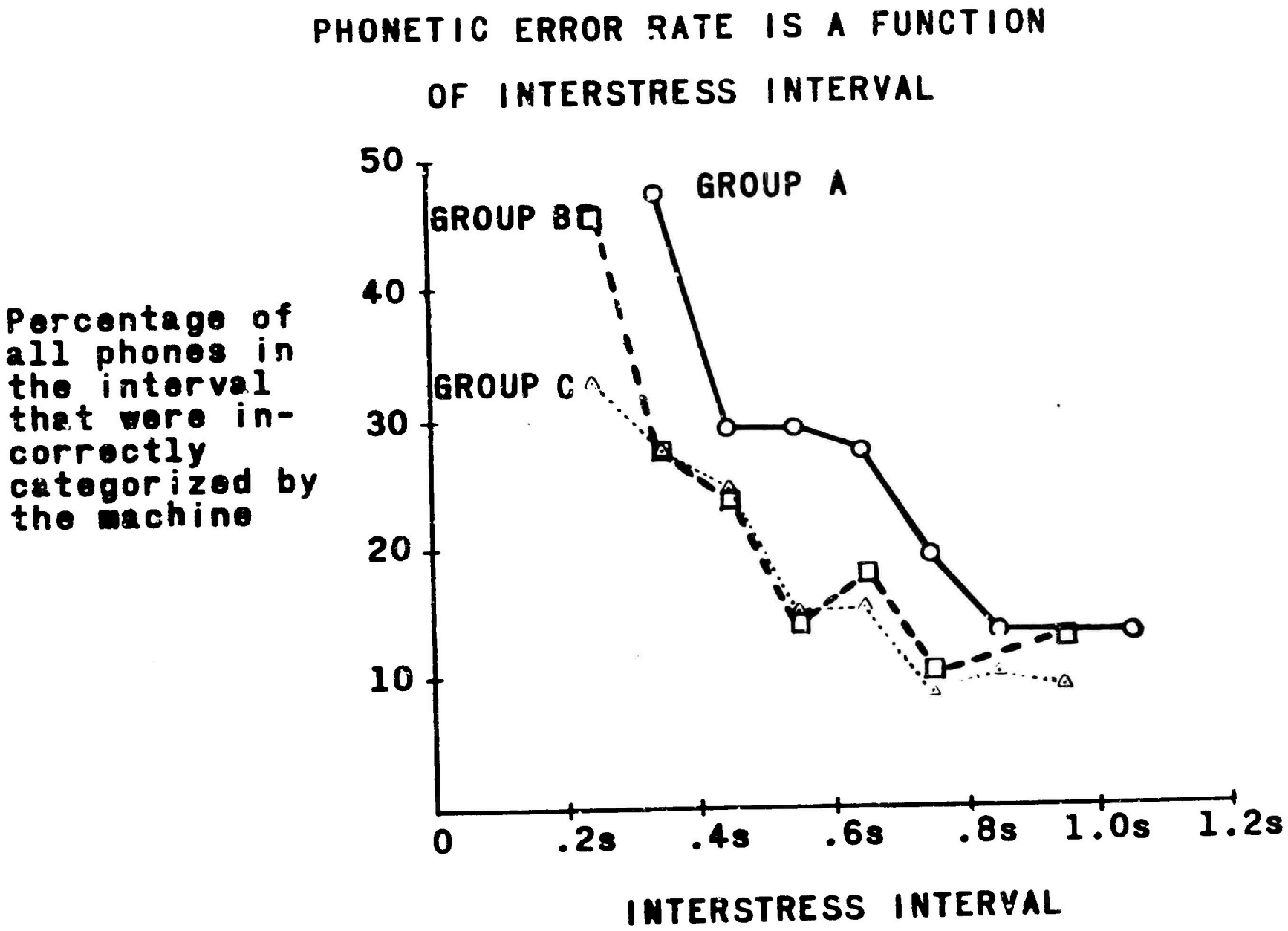


Figure 11. Error Rates for Automatic Phonetic Labelling were Higher When the Interstress Interval Was Shorter

more readily predictable from the interstress interval than from other measures. One other measure investigated was the average time per syllable (obtained by measuring the time between the onset of the first stressed vowel in a sentence and the onset of the last stressed vowel in the sentence, and dividing by the number of syllables included in that time). This time per syllable is a direct inverse of the number of syllables per second, which might have been expected to be a reasonable measure of speech rate. Table V shows a lack of substantial correlation between such a syllable-timing measure and the phonetic error rate. Another conceivable measure of speech rate, the average time per phone (obtained by dividing the time interval between the onsets of the first and last stressed vowels by the number of phones spanned), showed even poorer correlation with the phonetic error rates.

Table V. Correlation Coefficients of Phonetic Error Rates and Measures of Speech Rate

MEASURE OF SPEECH RATE	PHONETIC SEGEMENTATION METHOD		
	A	B	C
Interstress Interval	-.60	-.76	-.65
Mean Interstress Interval	-.55	-.74	-.70
Time Per Syllable	-.10	- 18	-.56
Time Per Phone	+.01	+.24	-.26

For each of the rate measures given in Table V, the same portions of the utterances were included; namely, all the speech between the onset of the first stressed vowel and the onset of the last stressed vowel. However, if a pause occurred in the midst of an utterance, the speech between the onsets of the stressed vowels on either side of the pause was excluded from the computations of all speech rates. Thus, interstress intervals which span pauses were excluded from consideration. Also, if there was a syllable perceived as stressed by one listener (but unstressed by the other listeners), all the speech between the immediately preceding and following stresses was excluded from consideration in these studies. In this way, all unusually long interstress intervals associated with either pauses or uncertainties in stress assignment were excluded from the measurements of speech rate and from the measurements of phonetic error rates.

The speech before the first stressed vowel, and after the last stressed vowel, was also obviously not included in the measurements of speech rates and error rates. (It is worth noting, however, that phonetic errors are particularly frequent in these utterance-initial and final positions. The reason for excluding such regions from consideration is that all measures of speech rate cannot be extended into those regions.)

It may actually prove to be very advantageous that a local measure of speech rate such as the interstress interval should prove to be more informative about phonetic errors than any average rate is. The local measure is easier to obtain, requires no additional averaging computations, and, most important of all, can be determined on-line in a real-time manner, without requiring a delay until the whole utterance is completed.

In summary, it appears that the duration of individual interstress intervals is the best measure (or one of the best measures) for relating speech rate to applicable phonetic or phonological rules. Knowing the interstress interval, one can then predict how likely it is that phonetic categorization errors may occur. Since it is the purpose of acoustic phonetic and phonological rules to account for such changes in phonological structure, we may expect that the interstress interval as a measure of speech rate may be used to predict what phonological rules may be suitable for applying at various points in spoken utterances.

Thus, our experiments have shown that the interstress interval can not only provide cues to phonological and syntactic boundaries; the interval may also be used to predict when phonological and acoustic phonetic rules should be applied, to determine underlying phonological forms and appropriate wording of the sentence.

4. CONCLUSIONS AND FURTHER STUDIES

4.1 Summary

Sperry Univac's previous series of experiments on prosodic structures have culminated in the delivery of computer programs for constituent boundary detection and stressed syllable location. In cooperation with BBN and SDC, we have just begun the vital task of integrating these prosodic analysis tools into the SDC and BBN speech understanding systems. Our recent experiments on timing cues to linguistic structure have broadened the scope of our prosodic studies, to include additional features of vowel and sonorant lengthening, interstress intervals, and pauses, as cues to phonological boundaries. Also, the interstress interval has been shown to be a good measure of speech rate which is suitable for predicting the phonetic variations that must be accounted for by acoustic phonetic and phonological rules.

In section 4.2, we will describe some further improvements that might be introduced into the programs for boundary detection and stressed syllable location. A plan for integrating various prosodic analysis programs under a prosodic executive routine is outlined in section 4.3. Other experimental work and applications to the speech understanding systems are outlined in section 4.4.

4.2 Improvements in the Prosodic Programs

The boundary program BOUND3, as currently implemented (and as it was distributed to ARPA SUR contractors), does not assign a boundary at the beginning of an utterance, nor at the end of the utterance unless a 35cs pause is detected. Presumably every speech understanding system will provide a decision about the positions of the beginning and ending of the utterance, so these delimiting brackets will be already available. Another utterance-final effect is that currently the confidences of the last constituents tend to be low, since no following constituents are there to add increments due to the magnitude of F_0 rise. This could be corrected by adding a fixed increment to the confidences of terminal constituents.

Another improvement that could be made in future implementations of the program could be to replace the 35 cs threshold on pauses by a user-defined variable (which could be adjusted for the rate of speech and the talker identity, etc.). Currently, the thresholds of F_0 rise or fall are given by the user, in his first data card. A value of 5 eighth tones has been used for each. Testing different values of these thresholds could help one optimize the performance of the program.

Another change may be to replace the current batch-mode of reporting results. Boundaries and summary tables are presently printed on computer listings, but in actual systems, the results should be used as arrays of data for use in subsequent programs.

In summary, BOUND3 appears to be a good program for detecting syntactic boundaries from F_0 contours. This program may be useful in speech understanding systems for any of several purposes: 1) determining some aspects of syntactic structure directly from acoustic data, independent of the segmental analysis of words within the sentence; 2) breaking the large sentence down into units of more manageable size (provided the user doesn't demand total independence across the detected boundaries, or strict placement of the boundaries); and 3) providing the initial data needed by the Sperry Univac stressed-syllable-location program. The latter point is of particular importance. Even if the user has no immediate plans to use prosodic cues to determine syntactic structure, he may wish to implement the boundary detection program if he plans to use the Sperry Univac stressed syllable location program, since STRESS works on the basis of "archetype contours" found in constituents between the boundaries detected by BOUND3.

There also are a number of details and general concepts in the STRESS program which might be improved. For example, the CHUNK subroutine needs to be improved, so that the threshold amounts of energy dip or rise are defined from next-to-extremal values rather than the extreme energy maxima or minima. Then, single points out of line with other energy values will not determine the value from which a 5dB-change is sought. Rather, the next-to-highest (or next-to-lowest) points would be used. If the extreme value is maintained for two or more time segments, then that extremal value would be the same as the next-to-extremal value. This slight change will improve the specification of beginning and end points of chunks. Also, we should allow a one-point dip below the threshold, along with surrounding values above the threshold value, to all be included within a chunk, so that the chunk duration will in some cases be better characterized than currently.

Another refinement in the definition of the beginning and end points of chunks might be to make the threshold of energy dips be some fraction of the distance from the maximum energy in the chunk down to the higher of the surrounding energy dips, as is done in the current total energy test. Then, the 5dB threshold amount could still be used to find substantial dips, but the end points of the chunk would be determined by

those points where energy is down a fraction of the amount in the dips surrounding a chunk. Such a fractional test for locating chunk end points might be better than a strict 5dB drop.

An important improvement will be to have a voicing test rule out all unvoiced regions as not being chunks before the list of chunks is given to the STRESS program. Then, one could remove the several instances of a voicing test as currently included in HEADER or OTHERS.

Ultimately, we would expect that segmental information could be used in the syllabification routine, to help break up some chunks which currently contain several syllables. For example, if prominent formant transitions (such as a major dip in F_3 followed by a rise in F_3) occurred in the middle of chunks, we might take them as cues to intervocalic sonorants (such as /r/'s), and segment the chunk in the area of those sonorants. Similarly, if we knew that in the midst of a long chunk the phonetic segmentation and categorization procedures found more than one vowel, with intervening nasals or such, we could divide the chunk into two or more syllables. Then, chunks so divided would not appear to be single stressed syllables, and some false alarms in stress location would be eliminated.

In the actual selection of stressed syllables, it may be worth considering whether the priority of F_0 rises as stress cues couldn't be weakened somewhat, to eliminate some current selections of the wrong chunks. Also, further conditions on the application of the total energy test in HEADER might be added, to prevent the few cases where the total-energy test reverses a prior decision which had correctly located a stressed HEAD.

One general area to consider in the refinement of HEADER and OTHERS involves improving the specification of search ranges for locating stresses near F_0 rises. Also, the choice of the correct chunk within such search regions might be improved by some means of collapsing the current redundant tests into fewer and simpler tests.

It is also worth considering whether the test for prepausal stresses shouldn't be made more restrictive, to reduce the number of prepausal unstressed syllables that are declared to be stressed. Also, a special test for locating the initial stress following a pause might be appropriate, since sometimes the first stress is associated with the peak F_0 in the constituent, and some other times, the first stress is properly associated with the rapidly-rising F_0 preceding that peak F_0 position.

It is possible that a more theoretical (or at least an empirically more satisfying) definition of the "archetype line" might be defined. The current procedure for selecting the TAIL is certainly crude. Also, a procedure might be needed to restrict the search for rises above the archetype line, so that rises right after the stressed HEAD are not considered. For example, we might require F_0 to drop below the archetype line first and then rise above it before a test for another stress is made. Also, maybe the chunk right after the HEAD should not be permitted to be another stress unless it has a definite F_0 rise and/or a quite long chunk duration.

Similarly, in the test for long chunks during long archetype lines (TTAIL - TTOP $\geq$ 100), we might insert some form of F_0 slope test, to reduce the likelihood of false "stresses" being found in long unstressed chunks within those regions.

There are a number of thresholds that are incorporated in the STRESS program that might be replaced by user-controlled variables. These include the 5dB energy dips, the selection of chunk end points, the energy ratios of .62, .64, and .75, the F_0 slope of two eighth tones per five segments, etc.

Ultimately, for efficiently handling continuous speech in a real-time mode, we may want to get the F_0 and energy data, and constituent boundary locations, constituent by constituent, so that the program can process on the order of one constituent at a time, go on the next one, and continue, without needing large data arrays.

Finally, we expect that it may be very valuable to assign confidences to the stressed syllable locations. We would use the confidences provided by the boundary detection program as a start in assigning confidences to stressed HEADS, then adjust them (and those found by OTHERS) based on the amount of F_0 rise in the chunk, the duration of the chunk, the energy values, and other factors such as the proximity to other stresses, rhythm, contextual comparisons, etc. Lower confidences could be assigned to stresses found by the "last resort" procedures in OTHERS, such as prepausal stresses, stresses within long archetype lines, and stresses found by short (two-segment) rises above the archetype line.

4.3 Plans for a Prosodic Executive Routine

Small input and output changes are needed in both BOUND3 and STRESS, to get data and results in the forms desired in specific systems and to have the programs

operate together. However, a much more vital change is ultimately needed if these and other prosodic routines are to be integrated into speech understanding systems. That is to incorporate the prosodic structure analyzers into a single "prosodic executive routine".

Figure 12 illustrates one way in which a number of prosodic analysis tools can be incorporated under one executive routine. The executive routine PROSOD calls BOUND4 (nearly identical to BOUND3), to detect constituent boundaries. Then the routine SYLLB2 is called to do a segmentation of the speech into syllables (voiced chunks corresponding to syllabic nuclei). A routine PAUSE would locate (and categorize) silent pauses. Routine STREST (which is to be a slight refinement of STRESS, with confidences assigned) would then determine stressed syllable locations based on these boundary detections, syllabification results, and (possibly) pause positions. A routine SHOW would display F_0 and energy versus time, with boundaries and maximum F_0 points marked, pauses shown, and stressed nuclei shown.

The time intervals between stresses, between syllables, and between boundaries would be used in a routine called RHYTHM, to determine: a) measures of speech rate, such as number of syllables per second, individual (and mean) time intervals between stresses, and mean time intervals between syntactic boundaries; b) rhythmic structures, such as long disjunctures, as cues to syntactic boundaries; and, c) gaps in rhythm, where another stressed syilable might have been expected.

We would then hope to have a routine GUIDE called by PROSOD, to provide the guidelines for efficient phonemic analysis, such as locating stressed vowels and the total stressed-syllable analysis region, where careful phonetic segmentation and classification could be attempted (cf. Lea, Medress, and Skinner, 1975).

Another routine called GUESS would make initial guesses as to the syntactic structure of the sentence, based on the number of syntactic boundaries and their positions, the locations of stresses, the disjunctures and pauses as cues to syntactic breaks, and the cues to sentence type, subordination, and coordination provided by the various prosodic patterns. The development of GUESS would be one of the primary outcomes out of the extensive studies to be conducted with our large speech data base.

4.4 Further Studies

In the current contract, the extensive sets of Phonetic Sentences (Lea, 1974b) and Prososyntactic Sentences (Lea, 1974c) will be recorded three times by each of

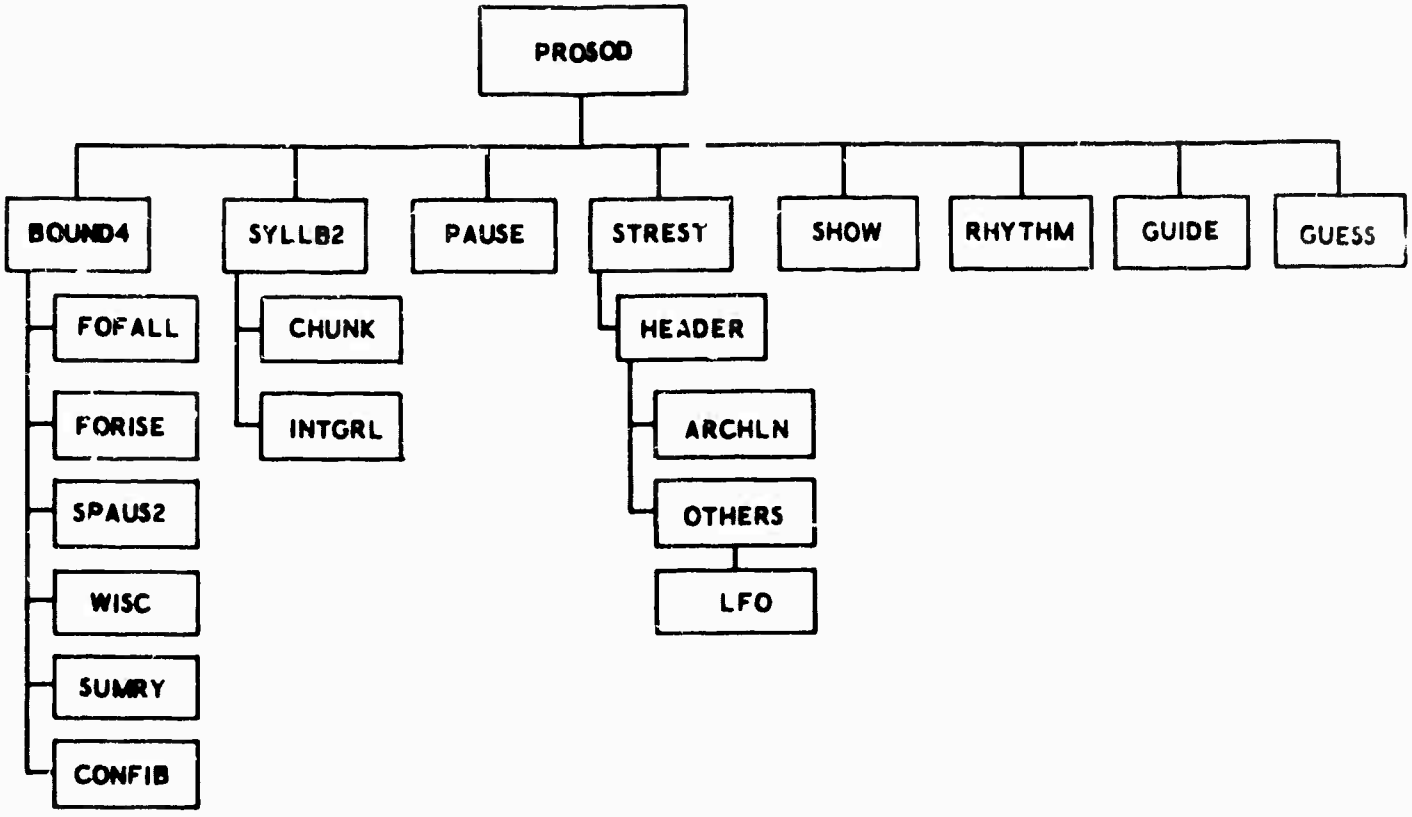


Figure 12. Structure of a Prosodic Executive Routine

three talkers. A subset of sentences which have minimal contrasts in the positions of stressed syllables within constituents will be processed through the acoustic parameterization and prosodic analysis programs. The prosodic patterns in such sentences will be analyzed to determine where the F_0-detected boundaries are positioned as the first stress in the following constituent is moved (e.g., from the first to the second to the third syllable).

In further work for ARPA, we plan to extend our initial timing studies, by investigating interstress intervals occurring in the various sentence structures included in the extensive Sperry Univac speech data base (Lea, 1974b, c). Then, such timing measures will be applied to specific tasks of aiding syntactic parsers and phonological rules components of speech understanding systems. These studies will involve some more of the designed sentences, and will also include more computer protocols being recorded for use with the SDC speech understanding system. Other experiments will

be conducted (using designed sentences and protocols) to determine whether prosodies can usefully provide cues to sentence type, contrastive syntactic bracketing, and subordination.

A major effort will be devoted to improving and extending our prosodic analysis programs, integrating them into existing speech understanding systems, and testing their effectiveness in aiding word matching and parsing procedures. Sperry Univac will also be involved in developing metrics and procedures for performance evaluation of speech understanding systems and their components.

5. REFERENCES

ALLEN, G. D. (1968), On Testing for Certain Stress-Timing Effects, Working Papers in Phonetics No. 10, University of California at Los Angeles, 47-59.

BLESSER, B., (1972), Speech Perception under Conditions of Spectral Transformation: I. Phonetic Characteristics, Jour. Speech and Hearing Res., 15, 5-41.

BOLINGER, D. (1958), A Theory of Pitch Accent in English. Word, vol. 15, p. 109.

HARRIS, M. S. and UMEDA, N., (1974), Effect of Speaking Mode on Temporal Factors in Speech: Vowel Duration, J. Acoust. Soc. of America, 56, 1016-1018.

KLATT, D. H. (1973), Interaction between Two Factors that Influence Vowel Duration, J. Acoust. Soc. America, 54, 1102-1104.

KLATT, D. H. (1974), On the Design of Speech Understanding Systems, Proc. Speech Comm. Seminar, ed. by G. Fant, Stockholm, 3, 277-289.

KLATT, D. H. (1975), Vowel Lengthening is Syntactically Determined in a Connected Discourse, submitted for publication.

KLOKER, D. R. (1975), Vowel and Sonorant Lengthening as Cues to Phonological Phrase Boundaries, presented at the 89th Meeting, Acoustical Society of America, Austin, Texas, April 8-11, 1975.

LEA, W. A. (1972), Intonational Cues to the Constituent Structure and Phonemics of Spoken English, Ph.D. Dissertation, School of Electrical Engineering, Purdue University.

LEA, W. A. (1973a), Syntactic Boundaries and Stress Patterns in Spoken English Texts, Univac Report No. PX 10146, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1973b), Segmental and Suprasegmental Influences on Fundamental Frequency Contours. In Consonant Types and Tone (L. Hyman, Ed.), Los Angeles: Univ. of Southern California Press, 15-70.

LEA, W. A. (1974a), Prosodic Aids to Speech Recognition: IV. A General Strategy for Prosodically-Guided Speech Understanding, Univac Report No. PX 10791, Sperry Univac, DSD, St. Paul, Minnesota

LEA, W. A. (1974b), Sentences for Controlled Testing of Acoustic Phonetic Components of Speech Understanding Systems, Univac Report No. PX 10952, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1974c), Sentences for Controlled Testing of Prosodic and Syntactic Components of Speech Understanding Systems, Univac Report No. PX 10953, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A. (1974d), Prosodic Aids to Speech Recognition: V. A Summary of Results to Date, Univac Report No. PX 11087, Sperry Univac DSD, St. Paul, Minnesota.

LEA, W. A., MEDRESS, M. F., and SKINNER, T. E. (1972), Use of Syntactic Segmentation and Stressed Syllable Location in Phonemic Recognition. Presented at the 84th Meeting, Acoustical Society of America, Miami Beach, Florida, Nov. 27-30, 1972.

LEA, W. A., MEDRESS, M.F., and SKINNER, T. E. (1973) Prosodic Aids to Speech Recognition III: Relationships between Stress and Phonemic Recognition Results, Univac Report No. PX 10430, Sperry Univac DSD, St. Paul, Minnesota.

LEA, WA. A., MEDRESS, M. F., and SKINNER, T.E. (1975), A Prosodically-Guided Speech Understanding Strategy. IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP 23, 30-38.

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

LEHISTE, I., The Timing of Utterances and Linguistic Boundaries, J. Acoust. Soc. Amer., 51, 2018-2024 (1972).

LIEBERMAN, P. (1967), Intonation, Perception, and Language, Cambridge: M.I.T. Press.

OLLER, D. K., The Effect of Position in Utterance on Speech Segment Duration in English, J. Acoust. Soc.America, 1235-1247 (1973).

PIKE, K. L. (1945), The Intonation of American English. Ann Arbor: University of Michigan.

6. APPENDIX

An eighth-tone scale for representing fundamental frequency values has been used in the prosodic analysis programs at Sperry Univac. This scale is based on the musical note A_2 being at 110 Hz, and each rise of an eighth tone being a frequency ratio of $\sqrt[48]{2.0}$, or 1.014545335. This yields 48 eighth tones in an octave.

Rather than use the formula in a computer, a table look-up is often more convenient. Table A-I is used. If F_0 in Hertz is between F_L and F_u inclusively, in Hertz, then assign F_0 to the eighth tone of frequency F_s, with the eighth tone number N_1. Note that $N_1 = 0$ for 69.3 Hz. This was originally set because in our earliest studies we didn't expect to need F_0 values below about 70 Hz. However, later work required the use of negative numbers, to cover F_0 values down to 32 Hz (eighth tone number $N_1 = -53$). The next lower value, $N_1 = -54$, was then used for unvoiced segments.

TABLE A-I. EIGHTH TONE SCALE FOR REPRESENTING F_0 VALUES

If F_0 in Hertz is between F_ℓ and F_u, inclusively, then assign the standard eighth tone value F_s which is assigned the corresponding eighth tone number N_1.

F_ℓ	F_u	F_s	N_1	F_ℓ	F_u	F_s	N_1	F_ℓ	F_u	F_s	N_1	F_ℓ	F_u	F_s	N_1
0	0	0	-54	81	81	.23	11	162	163	162.45	59	323	327	324.90	107
32	32	32.23	-53	82	83	82.41	12	164	166	164.81	60	328	332	329.63	108
33	33	33.18	-51	84	84	83.61	13	167	168	167.21	61	333	336	334.42	109
34	34	34.15	-49	85	85	84.82	14	169	170	169.64	62	337	341	339.29	110
35	35	35.15	-47	86	86	86.06	15	171	173	172.11	63	342	346	344.22	111
36	36	36.18	-45	87	88	87.31	16	174	175	174.61	64	347	350	349.23	112
37	37	37.24	-43	89	89	88.58	17	176	178	177.15	65	352	356	354.31	113
38	38	37.78	-42	90	90	89.87	18	179	181	179.73	66	357	362	359.46	114
39	39	38.89	-40	91	91	91.17	19	182	183	182.34	67	363	367	364.69	115
40	40	40.03	-38	92	93	92.50	20	184	186	185.00	68	368	372	370.00	116
41	41	41.20	-36	94	94	93.84	21	187	189	187.69	69	373	378	375.38	117
42	42	41.80	-35	95	95	95.21	22	190	191	190.42	70	379	383	380.84	118
43	43	43.03	-33	96	97	96.59	23	192	194	193.19	71	384	389	386.38	119
44	44	44.29	-31	98	98	98.00	24	195	197	196.00	72	390	394	392.00	120
45	45	44.93	-30	99	100	99.42	25	198	200	198.85	73	395	400	397.70	121
46	46	46.25	-28	101	101	100.87	26	201	203	201.74	74	401	406	403.48	122
47	47	46.92	-27	102	103	102.34	27	204	206	204.68	75	407	412	409.35	123
48	48	48.30	-25	104	104	103.83	28	207	209	207.65	76	413	418	415.30	124
49	49	49.00	-24	105	106	105.34	29	210	212	210.67	77	419	424	421.35	125
50	50	49.71	-23	107	107	106.87	30	213	215	213.74	78	425	430	427.47	126
51	51	51.17	-21	108	109	108.42	31	216	218	216.85	79	431	436	433.69	127
52	52	51.91	-20	110	110	110.00	32	219	221	220.00	80	437	443	440.00	128
53	53	52.67	-19	111	112	111.60	33	222	224	223.20	81	444	449	446.40	129
54	54	54.21	-17	113	114	113.22	34	225	228	226.45	82	450	455	452.90	130
55	55	55.00	-16	115	115	114.87	35	229	231	229.74	83	456	462	459.48	131
56	56	55.80	-15	116	117	116.54	36	232	234	233.08	84	463	469	466.16	132
57	57	56.61	-14	118	119	118.24	37	235	238	236.47	85	470	476	472.94	133
58	58	58.27	-12	120	120	119.96	38	239	241	239.91	86	471	483	479.82	134
59	59	59.12	-11	121	122	121.70	39	242	245	243.40	87	484	490	486.80	135
60	60	59.98	-10	123	124	123.47	40	246	248	246.94	88	491	497	493.88	136
61	61	60.85	-9	125	126	125.27	41	249	252	250.53	89	498	504	501.07	137
62	62	61.74	-8	127	128	127.09	42	253	256	254.18	90	505	512	508.36	138
63	63	62.63	-7	129	129	128.94	43	257	259	257.87	91	513	519	515.75	139
64	64	63.54	-6	130	131	130.81	44	260	263	261.63	92	520	527	523.25	140
65	65	65.41	-4	132	133	132.72	45	264	267	265.43	93	528	534	530.86	141
66	66	66.36	-3	134	135	134.65	46	268	271	269.29	94	535	542	538.58	142
67	67	67.32	-2	136	137	136.60	47	272	275	273.21	95	543	550	546.42	143
68	68	68.30	-1	138	139	138.59	48	276	274	277.18	96	551	558	554.37	144
69	69	69.30	0	140	141	140.61	49	280	283	281.21	97	559	-	562.43	145
70	70	70.30	1	142	143	142.65	50	284	287	285.30	98				
71	71	71.33	2	144	145	144.73	51	288	291	289.45	99				
72	72	72.36	3	146	147	146.83	52	292	295	293.66	100				
73	74	73.42	4	148	150	148.97	53	296	300	297.94	101				
75	75	74.38	5	151	152	151.13	54	301	304	302.27	102				
76	76	75.57	6	153	154	153.33	55	305	308	306.67	103				
77	77	76.67	7	155	156	155.56	56	309	313	311.13	104				
78	78	77.78	8	157	158	157.83	57	314	317	315.65	105				
79	79	78.91	9	159	161	160.12	58	318	322	320.24	106				
80	80	80.06	10												