DOCUMENT RESUME

ED 088 286 FL 005 218

AUTHOR	Lea, Wayne A.
TITLE	An Algorithm for Locating Stressed Syllables in Continuous Speech.
PUB DATE	31 Oct 73
NOTE	17p.; Paper presented at the Meeting of the Acoustical Society of America (86th, Los Angeles, California, October 3, 1973)
EDRS PRICE	MF-$0.75 HC-$1.50
DESCRIPTORS	Acoustic Phonetics; *Algorithms; Auditory Discrimination; Componential Analysis; Computers; *Connected Discourse; Consonants; Data Analysis; *Distinctive Features; Graphs; Input Output; Language Patterns; Language Research; Man Machine Systems; Phonological Units; *Phonology; Phrase Structure; *Syllables; Tables (Data)

ABSTRACT

Local increases in fundamental frequency (Fo) and large integrals of energy in the syllabic nucleus are known to be among the best acoustical correlates of stress. Major syntactic constituents have been shown to have archetype rapid-rise-then-gradual-fall Fo contours, with the rise into the maximum Fo often associated with the first stressed syllable in the constituent. An automatic procedure for detecting constituent boundaries and maximum Fo positions in constituents, and sonorant energy and Fo functions, provided input data for an algorithm for locating stressed syllables. The first stressed syllable of a constituent was associated with a high-energy-integral portion near the rising Fo into maximum Fo position. Other stressed syllables were associated with high-energy-integral portions near local increases in Fo above a steadily-falling "archetype line" from the maximum Fo position to the end of the constituent. For over 400 seconds of speech, including written texts, questions, commands, and declarations for man-machine interaction, over 85% of all syllables perceived as stressed by a panel of listeners were correctly located. (Author/DD)

AN ALGORITHM FOR LOCATING STRESSED SYLLABLES IN CONTINUOUS SPEECH

Wayne A. Lea
Sperry Univac DSD
P.O. Box 3525
St. Paul, Minn. 55165

ABSTRACT

Local increases in fundamental frequency (F_0) and large integrals of energy in the syllabic nucleus are known to be among the best acoustical correlates of stress. Major syntactic constituents have been shown to have archetype rapid-rise-then-gradual-fall F_0 contours, with the rise into the maximum F_0 often associated with the first stressed syllable in the constituent. An automatic procedure for detecting constituent boundaries and maximum F_0 positions in constituents (Lea, W. A. (1973), An Approach to Syntactic Recognition without Phonemics, IEEE Trans. Audio and Electroacoustics, AU-21, No. 3), and sonorant energy and F_0 functions, provided input data for an algorithm for locating stressed syllables. The first stressed syllable of a constituent was associated with a high-energy-integral portion near the rising F_0 into maximum F_0 position. Other stressed syllables were associated with high-energy-integral portions near local increases in F_0 above a steadily-falling "archetype line" from the maximum F_0 position to the end of the constituent. For over 400 seconds of speech, including written texts, and questions, commands, and declarations for man-machine interaction (involving fifteen talkers), over 85% of all syllables perceived as stressed by a panel of listeners were correctly located.

Presented at the 86th Meeting, Acoustical Society of America, Los Angeles, California, October 31, 1973. 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.

AN ALGORITHM FOR LOCATING STRESSED SYLLABLES IN CONTINUOUS SPEECH

Wayne A. Lea

An algorithm for locating stressed syllables from prosodic features of energy and fundamental frequency has been devised. It is based on local increases in fundamental frequency, and large integrals of energy within the syllabic nucleus, being the most reliable acoustic correlates of stress. This algorithm also incorporates adjustments based on the most common ("archetype") fundamental frequency contours within the grammatical phrases and clauses of connected speech.

Connected speech texts whose stress patterns were studied included a paragraph of the Rainbow Script read by six talkers, a paragraph composed of only monosyllabic words ("Monosyllabic Script") read by two talkers, and 31 spontaneous sentences intended for man-computer interaction, which had been recorded by nine talkers involved in the ARPA Speech Understanding Research Program. In a companion study reported on in another paper at this meeting, a panel of listeners repeatedly heard these spoken texts until they could provide judgments as to which syllables were stressed, unstressed, or reduced.

The spoken scripts were processed through an autocorrelation algorithm for fundamental frequency tracking (or "pitch" tracking), and through an algorithm which provided a so-called "sonorant" energy function, which gives the speech energy within the frequency range of 60 Hz to 3000 Hz. This sonorant energy function should give high energy values within sonorant syllabic nuclei, while giving lower values during obstruents.

The first slide shows a stylized plot of fundamental frequency, on a logarithmic or eighth-tone scale, and a corresponding plot of sonorant energy on a dB scale. The algorithm then operates on this data as follows. First, as the next slide shows, the connected speech is segmented into sentences and major grammatical constituents by an algorithm for detecting phrase boundaries at the bottoms of substantial fall-rise "valleys" in fundamental frequency contours (Lea, 1971, 1972, 1973). The increasing fundamental frequency near the beginning of each constituent is assumed to be attributable to the first stressed syllable or "HEAD" of the constituent, as shown on the next slide. A portion of the speech which is high in energy with increasing fundamental frequency values, and which is bounded by points where the energy dips 5 dB or more, is asserted to be the stressed nucleus of this HEAD syllable. This is shown by the blue-tinted portions in this slide. Previous studies have shown that this stress-induced initial rise in fundamental frequency in a constituent is usually followed by a gradual fall in fundamental frequency, which may be approximated by a straight line on the logarithmic frequency scale. As shown in the next slide, the "archetype line" steadily drops in eighth tone values from the maximum fundamental frequency in the constituent down to the low value at the end of the constituent. Other stressed syllables in the constituent are expected to be accompanied by local increases in fundamental frequency - increases which make the fundamental frequency contour locally rise above the archetype line. Thus, even though fundamental frequency may not be rising absolutely at such stressed syllables, the fact that it is not falling at its usual rate can be a cue to the presence of a stressed syllable. The stressed syllable is again located within a high-energy region bounded by 5 dB dips in energy, as shown by the new yellow-tinted portions on the slide.

Detailed descriptions of this algorithm are available in published reports (Lea, 1973; Lea, Medress, and Skinner, 1973). The next slide shows the overall comparison between the algorithmically located stressed syllables and the listeners' perceptions of stressed syllables. For each text, and with results pooled for talkers, the table here gives the percentages of all syllables perceived as stressed (by two or more listeners) that the algorithm correctly located within the high-energy portions of speech. Occasionally the algorithm located a stretch of speech that did not enclose any syllable perceived as stressed by the listeners. Dividing the number of such <u>false</u> locations by the total number of algorithmically located portions gives the percentage of all locations that were false.

While scores varied somewhat from text to text and talker to talker, the overall average of 86% correct location of stressed syllables is very encouraging. Scores for the Rainbow Script read by six talkers ranged from 78% to 98%. Results for only two talkers reading the Rainbow Script are shown pooled here, for ease of direct comparison with results by the same two talkers reading the Monosyllabic Script. The Monosyllabic Script, with its fewer reduced syllables and more prominant stresses on monosyllabic content words, yielded quite high scores. The spontaneous ARPA Sentences, which were more monotone and which gave some difficulties to the constituent boundary detection algorithm, showed lower stressed syllable location scores. False locations resulted from falsely detected syntactic constituent boundaries and "borderline" cases of syllables perceived as stressed by at least one individual listener. Some of the failures to locate stressed syllables resulted from lack of fundamental frequency increases on some stressed syllables. A few failures resulted from more than one stressed syllable being within the initial portion of the constituent that has increasing fundamental frequency. The ultimate <u>use</u> of a stressed syllable location algorithm will determine whether false alarms or failures to locate stressed syllables are the least desirable errors.

To further evaluate the effectiveness of this archetype contour algorithm for locating stressed syllables, these results were compared with results in stressed syllable location by other procedures. The next slide shows one simple procedure which finds all dips and peaks in the sonorant energy function and delimits syllabic nuclei as all contiguous points within 5 dB of the maximum intensity value in each high-intensity "chunk" or syllable. Then, those chunks (or syllabic nuclei) that have a minimum duration of 100 ms are declared to be stressed.

Another simple subroutine, shown in the next slide, locates all portions of speech where, for 100 ms or longer, fundamental frequency does not decrease more than one eighth tone per ten milliseconds (this is sort of a relaxed form of a process of finding regions where fundamental frequency is steadily rising, or at least not falling rapidly).

The next slide shows that the location of stressed syllables from durations of high-intensity chunks works surprisingly well in read texts with sharply contrasting stress levels, such as the Monosyllabic Script, but it is not as effective in more complicated read texts such as the Rainbow Script or in spontaneous speech such as the ARPA Sentences. Lowest percentages of correct location and highest percentages of false alarms occur for the spontaneous ARPA sentences. The next slide shows that regions of increasing fundamental frequency are also less reliably related to stressed syllables in such sentences, and generally give poorer performance even in the Monosyllabic Script. The archetype-contour algorithm obviously performs better than either of these two simpler algoirthms, particulary for spontaneous speech. The next slide summarizes relative performance of the algorithms, showing that about 10% more stressed syllables are correctly located and about 10% fewer false alarms occur for the archetype-contour algorithm.

The last slide shows how stressed syllable location by the algorithms is affected by the type of sentence spoken (for the ARPA Sentences). For each algorithm, false alarms (shown within the orange boxes) are most frequent in yes/no questions. As shown within the yellow bands the <u>lowest</u> correct location score from chunk durations occurs in yes/no questions, while the <u>highest</u> correct location score from increases in fundamental frequency occurs in yes/no questions. This suggests the value of combining the two types of cues to improve success in stressed syllable location, such as is done in the archetype-contour algorithm.

In general, it is apparent that fairly accurate procedures are available for locating stressed syllables in continuous speech, particularly for read texts with sharp stress contrasts. Even the simplest procedures can locate on the order of 75% or more of the stressed syllables, but complex algorithms seem to be approaching 95% location with on the order of 20% false alarms. Further improvements now being implemented include other combinations of energy and fundamental frequency cues, and the incorporation of confidence measures to assess just how sure each algorithm is that each portion of speech is or is not a stressed syllable. Further studies will be conducted using designed speech texts which isolate effects that sentence type, constituent structure, different lexical insertions, and phonetic content have on the location of stressed syllables.

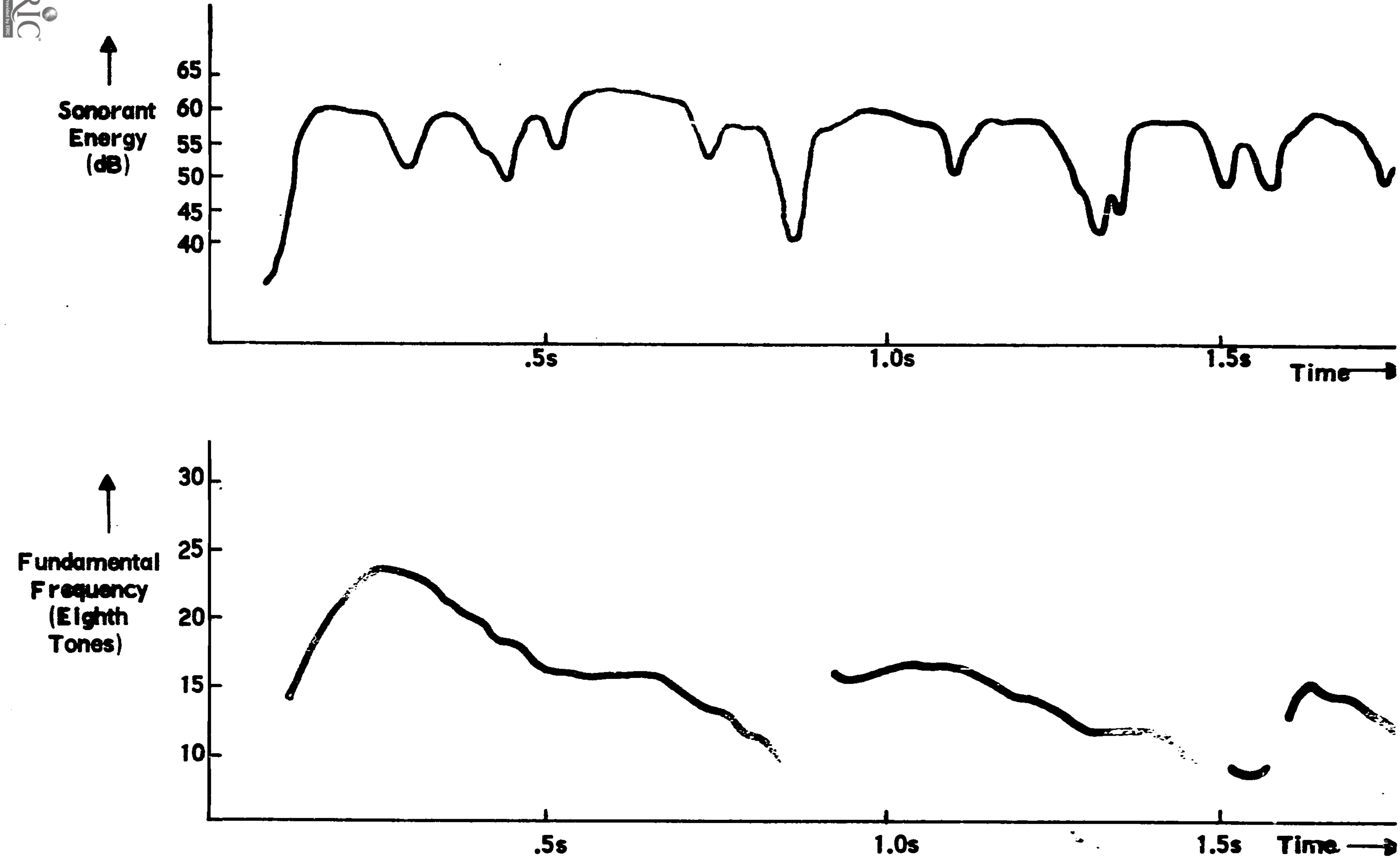
REFERENCES

LEA, W. A. (1971), Automatic Detection of Constituent Boundaries in Spoken English, J. Acoust. Soc. of America, vol. 50, 116 (A), July, 1971.

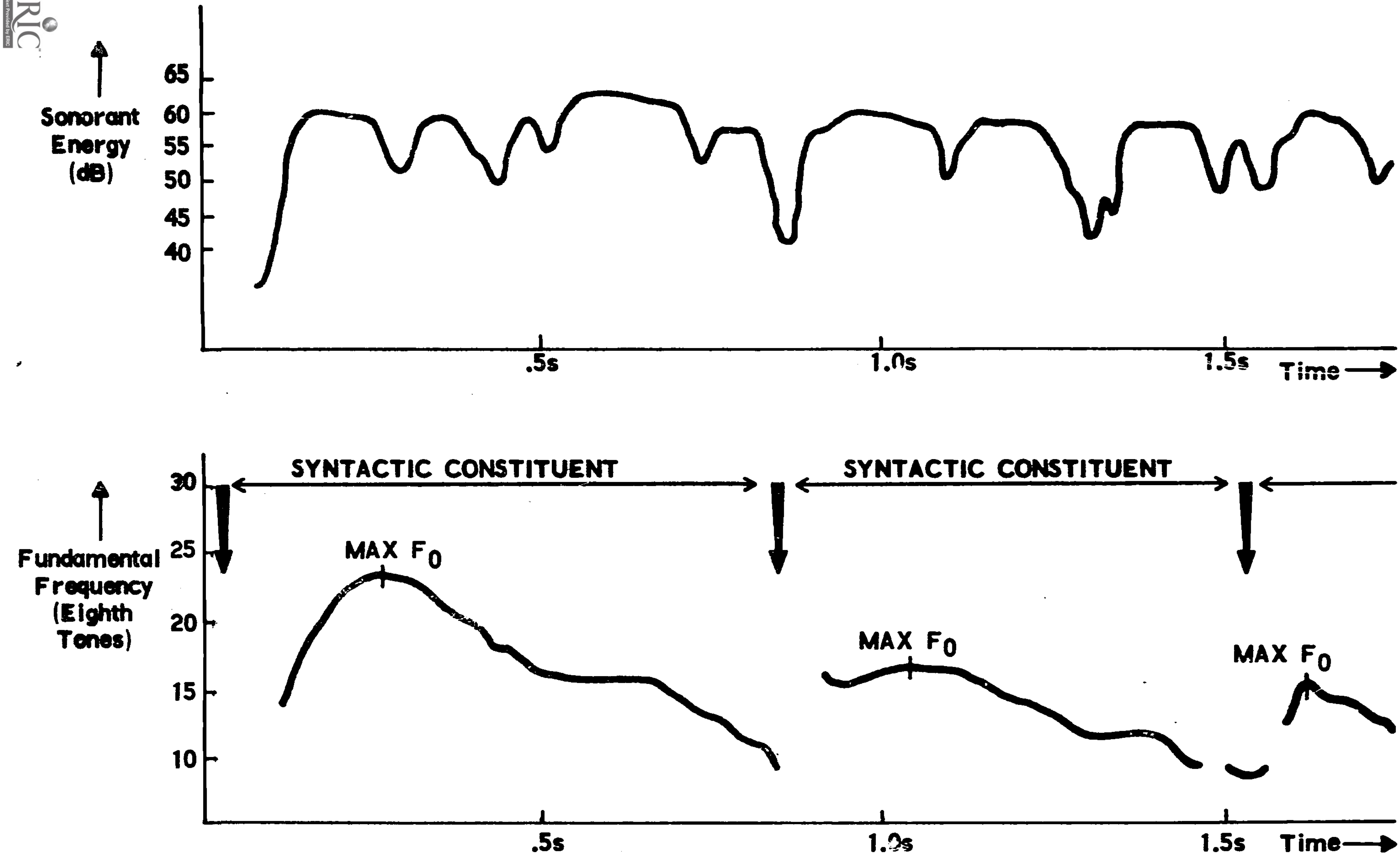
LEA, W. A. (1972), Intonational Cues to the Constituent Structure and Phonemics of Spoken English, Ph.D. Thesis, School of E. E., Purdue University.

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

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, Univac Park, St. Paul, Minnesota.

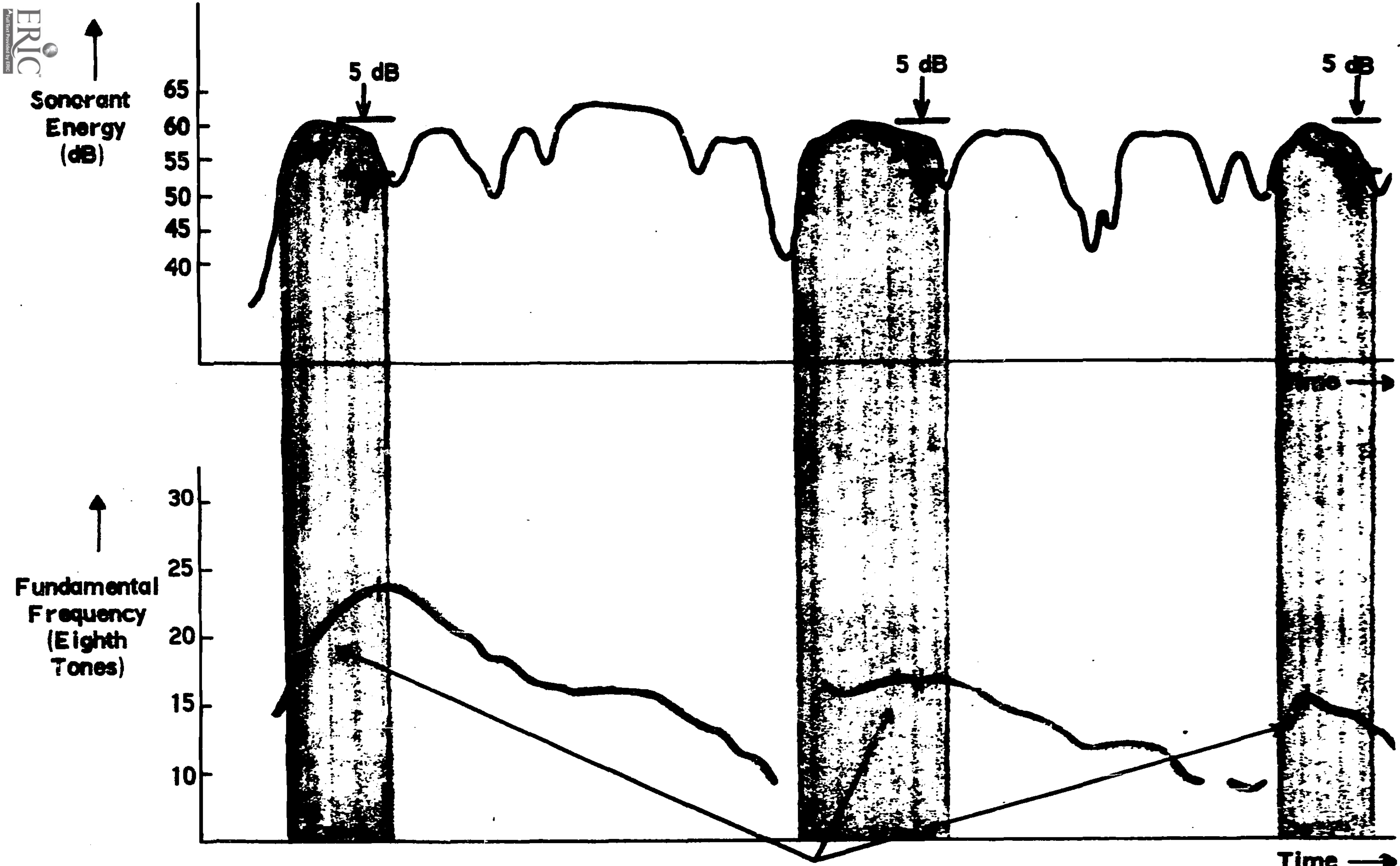

Sonorant Energy (dB)
65
60
55
50
45
40
.5s
1.0s
1.5s
Time
Fundamental Frequency (Eighth Tones)
30
25
20
15
10
.5s
1.0s
1.5s
Time

Slide
1



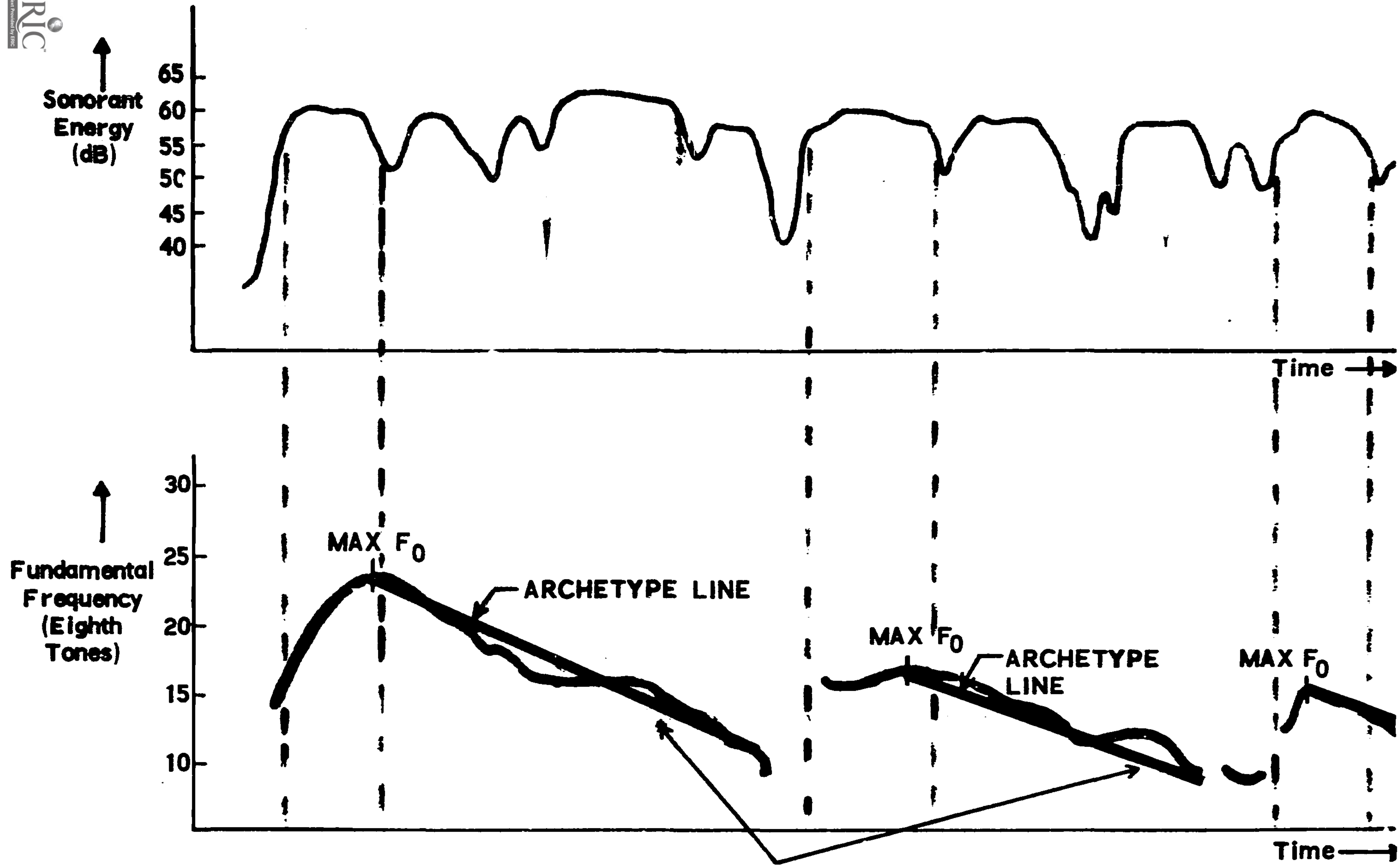
Sonorant Energy (dB)
65
60
55
50
45
40
.5s
1.0s
1.5s
Time
Fundamental Frequency (Eighth Tones)
30
25
20
15
10
SYNTACTIC CONSTITUENT
SYNTACTIC CONSTITUENT
MAX F0
MAX F0
MAX F0
.5s
1.0s
1.5s
Time

Slide
2



Substantial increases in fundamental frequency are cues to nearby stressed "Heads" in the beginnings of syntactic constituents. The stressed head syllables are then associated with nearby high-energy "chunks".

Slide
3



Local increases above the Archetype lines indicate other stressed syllables in the constituents. These are then located within nearby High-Energy "Chunks".

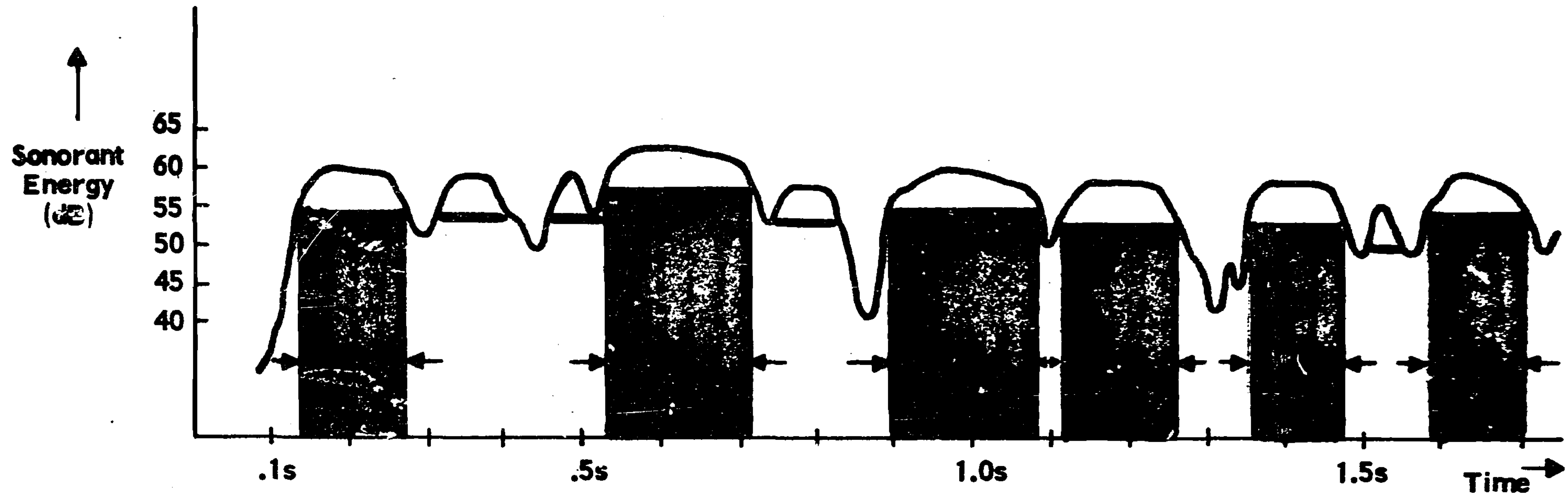
Slide
4

PERCENTAGES OF STRESSED SYLLABLE LOCATIONS WITH ARCHETYPE-CONTOUR ALGORITHM

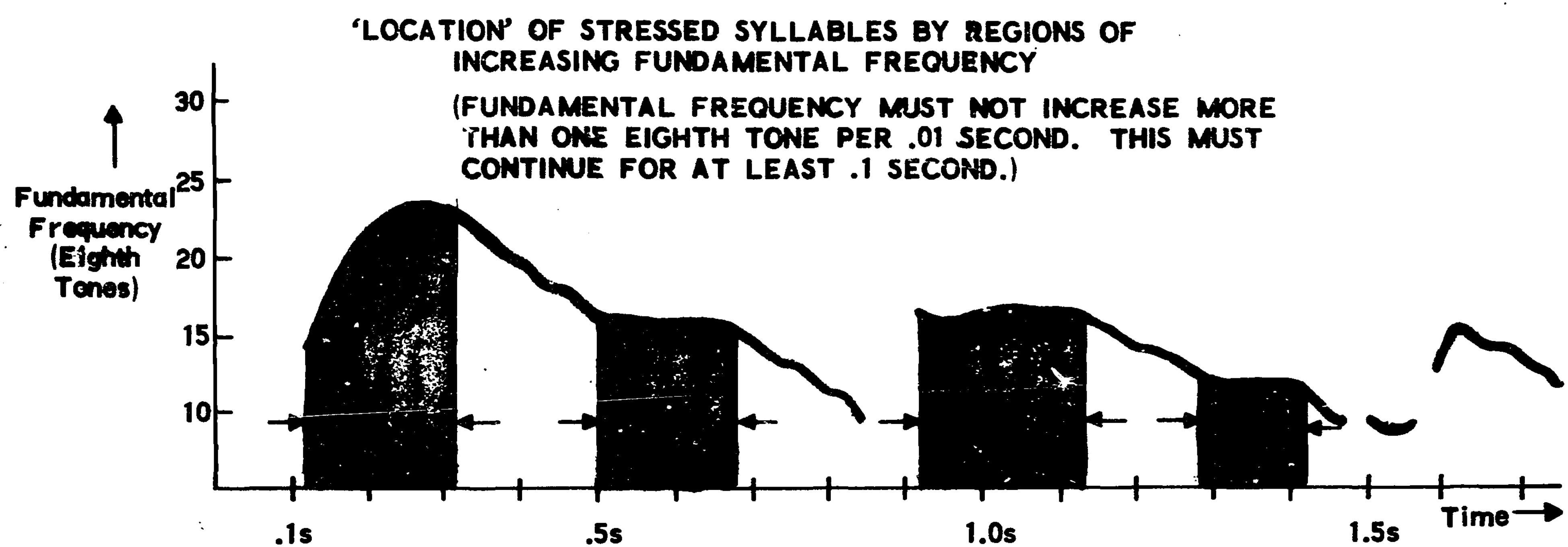
	RAINBOW (Only 2 Talkers)	MONOSYLLABIC (2 Talkers)	ARPA (9 Talkers)
CORRECT (PERCENTAGE OF THOSE SYLLABLES PERCEIVED AS STRESSED THAT WERE CORRECTLY LOCATED)	91%	93%	86%
FALSE ALARMS (PERCENTAGE OF ALL LOCATIONS THAT DID NOT INCLUDE A SYLLABLE PERCEIVED AS STRESSED)	16%	22%	23%

LOCATION OF STRESSED SYLLABLES BY HIGH-ENERGY CHUNKS OF LONG DURATION

(ENERGY MUST REMAIN WITHIN 5 DB OF PEAK FOR .1 SECOND)



Slide 6


'LOCATION' OF STRESSED SYLLABLES BY REGIONS OF INCREASING FUNDAMENTAL FREQUENCY
(FUNDAMENTAL FREQUENCY MUST NOT INCREASE MORE THAN ONE EIGHTH TONE PER .01 SECOND. THIS MUST CONTINUE FOR AT LEAST .1 SECOND.)
Fundamental Frequency (Eighth Tones)
30
25
20
15
10
.1s
.5s
1.0s
1.5s
Time

Slide
7

PERCENTAGES OF STRESSED SYLLABLE LOCATIONS FROM DURATIONS OF HIGH-ENERGY "CHUNKS"

	RAINBOW (2 Talkers)	MONOSYLLABIC (2 Talkers)	ARPA (9 Talkers)
CORRECT	80%	94%	76%
FALSE	25%	25%	38%

Slide
8

PERCENTAGES OF STRESSED SYLLABLE LOCATIONS FROM INCREASES IN FUNDAMENTAL FREQUENCY

	RAINBOW	MONOSYLLABIC	ARPA
CORRECT	79%	84%	73%
FALSE	22%	23%	26%

Slide
9

SUMMARY OF STRESSED SYLLABLE LOCATION BY THREE ALGORITHMS

	CORRECT	FALSE
ARCHETYPE ALGORITHM	90%	18%
DURATION OF HIGH-ENERGY CHUNKS	84%	28%
INCREASES IN FUNDAMENTAL FREQUENCY	77%	24%

Slide 10

EFFECTS OF SENTENCE TYPE ON STRESSED SYLLABLE LOCATIONS

		DECLARATIVES	COMMANDS	WH QUESTIONS	YES/NO QUESTIONS
ARCHETYPE ALGORITHM	Correct	88%	81%	87%	93%
	False	13%	23%	9%	30%
DURATIONS OF HIGH-ENERGY CHUNKS	Correct	79%	74%	83%	66%
	False	29%	39%	37%	49%
INCREASES IN FUNDAMENTAL FREQUENCY	Correct	72%	71%	70%	82%
	False	21%	23%	24%	38%