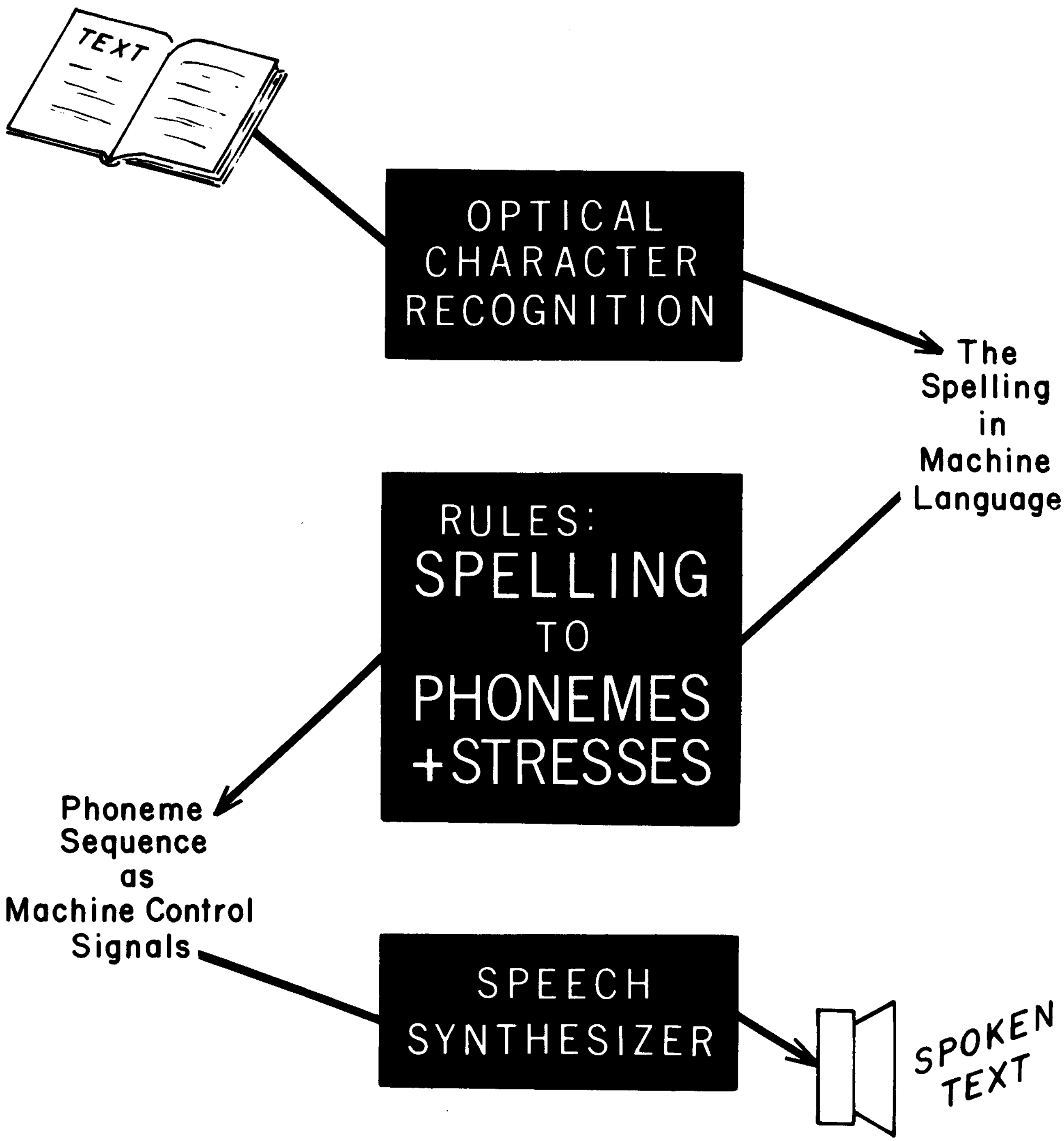
Print Reading by Machine



SLIDE A

Rules for Word Stress Analysis for Conversion of Print to Synthetic Speech*

Jane H. Gaitenby

Haskins Laboratories, New York City

Reading as done by a machine should, ideally, be comparable to human performance in the task of reading nonsense words aloud. Word pronunciation must be based on the spelling of the word, and sentence intonation will depend entirely on punctuation. Notwithstanding these inhuman handicaps, a reading machine is intended to produce an intelligible spoken text -- for human listeners -- and for blind listeners in particular.

Slide A shows the reading process as it is to be done by one type of machine. The three black boxes represent its main operations. An optical character recognizer transforms print to machine language; then, in the center, the spelled word becomes the phonetic word, by rule. (The subject of this paper applies here.) Finally, the phonetic symbols become commands for synthesizing speech. This last step involves speech synthesis by rule, which is a rapidly-advancing art as demonstrated recently at Cambridge, Mass. by Drs. Mattingly and Haggard. They have devised programs for the synthesis of American English and British English, using the Haskins computer and synthesizer. The form of computer input used by both Mattingly and Haggard was a man-made phonetic transcription.

We now want to program the machine to make its own phonetic judgments of the textual input, without human help. To do this, we must write clear-cut rules for the conversion of printed letters to speech-like phones. (We are not alone in this attempt,

* Paper given at the Seventy-fourth Meeting of the Acoustical Society of America, Miami Beach, Fla., 15 November 1967. To be published in the New Outlook for the Blind, Autumn 1968.

SLIDE B

SAMPLE NONSENSE WORDS IN TESTS

TEST I	TEST II	TEST III	TEST IV
snacial	snacially	snacene	snaciocrity
erchon	erchone	chonation	erchonibund
ciet	cieter	precietly	precietographic
busion	busioned	imbusion	busiopathy
adgiculy	adgicule	adgiculation	adgiculiognomy

but there are problems enough in English spelling to go around.)

As a start, we set up a test of nonsense words for humans to read, to observe their approach to a reading problem of comparable difficulty for man and machine. The general purpose was to determine the kinds of pronunciation cues -- and the order of cues -- employed by people when reading meaningless syllables. The nonsense items resembled English words: normal letter sequences were used, and many of the words contained familiar affixes. Slide B shows samples of the words used.

The specific purpose of the test was to see how people stress unknown words -- in particular, to pinpoint the role of affixes in the assignment of word stress.

Stress is a speech property, sometimes called "prominence," that can be perceived in any spoken syllables. There are numerous degrees of stress (all of them relative). In describing speech, linguists designate three -- or more often four -- degrees of stress.

"Word stress" refers to the gross rhythmic contour that accompanies the fixed syllable sequence of any polysyllabic word. Acoustically, stress depends on several interacting parameters: duration, intensity, fundamental frequency and even spectral distribution. Thus a strongly-stressed syllable, using all these cues, will be unusually high -- or low -- in voice pitch, long, loud, or it may be set off by silence from its neighbors.

Returning briefly to the nonsense test: a total of 150 words including those in Slide B were presented in four test sessions to 12 adult readers, individually. Before listing the assumptions that were tested, let me say that all were substantially confirmed.

"Stable" Affixes and Stress Prediction

Suffix Examples
(Predict stress from suffix)

INERTIA

DEMONIC

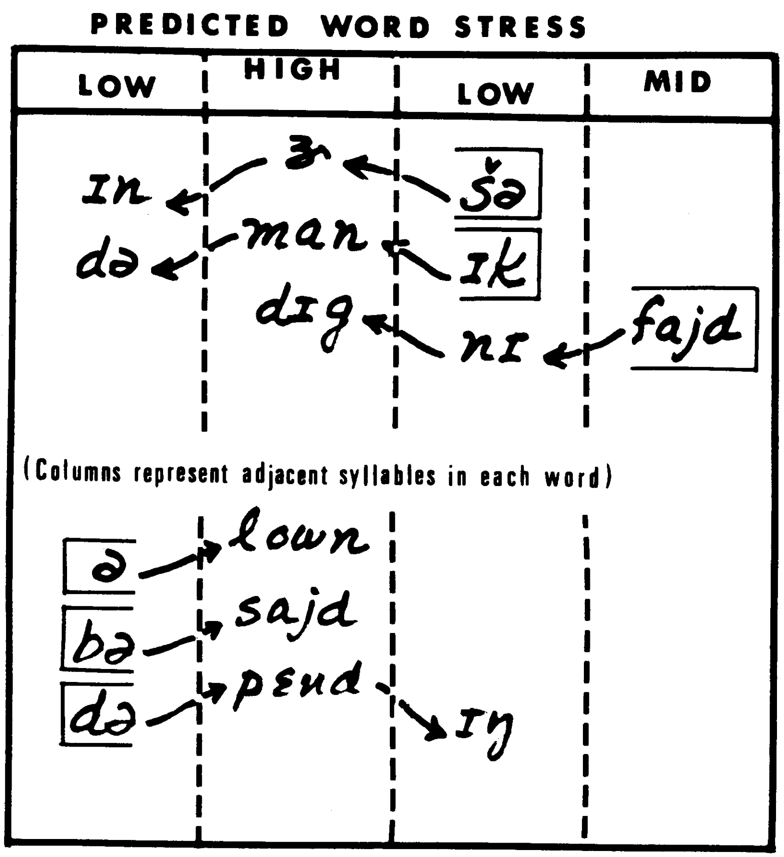
DIGNIFIED

Prefix Examples
(Predict stress from prefix)

ALONE

BESIDE

DEPENDING



SLIDE C

The assumptions were:

1. Some affixes are highly stable in stress properties, and provide the stress framework for all (or nearly all) words to which they are attached. Slide C shows three stable prefixes at the lower left.)
2. If no affix is present to force a different pattern, stress is strong on the first syllable of two and three syllable words.
3. Successive syllables in a word tend to be pronounced with alternating high and low stresses -- unless either the root, or the affix region, is compound.
4. Stress alternates backward from a stable suffix (look at Slide C again), or forward from a stable prefix. When both a suffix and prefix are present and stable, the suffix determines the word stress.
5. The stress pattern attributed to a word <u>limits</u> the phonetic quality given to its syllables, the <u>vowels</u> being most noticeably affected. For example,
   " <u>in</u>valid " with inverted alternating stress becomes
   " in<u>val</u>id "

In sum, these assumptions, which now serve as <u>working</u> rules, are based on the great frequency of stable affixes in actual texts, and on the tendency of stress to alternate in the syllables of a word. Further, the habitual stress level of a stable affix allows us to extrapolate the stress of the rest of the word.

If we combine stress probability with the observation that vowel color reflects the level of syllable stress quite dependably, we have a foothold on print to speech conversion. Slide D shows the single vowel letters, and three phonetic states probable for

SOUND CORRESPONDENCES FOR SINGLE VOWELS, BY SYLLABLE STRESS TYPE

Letter	Stress & Sound HIGH	MID	LOW
A	æ	æ	ə*
E	ɛ	ɛ	ə
I	I	I	I
O	a	a	ə
U (INIT.: ju)	ʌ	ʌ	jə

Letter	Stress & Sound HIGH	MID	LOW
A+cons+E	ej	ej	I
E " "	i	i	I
I " "	aj	aj	I
O " "	oʷ	oʷ	ə
U " "	ju	ju	jə

* "ALABAMA" > \| æ \| lə \| bæ \| mə
MID LOW HIGH LOW

SLIDE D

SLIDE E

SYLLABIFICATION RULES

V|V*	as in	¦c v¦c v|v c ¦v a¦c u|u m
V|C	" "	¦c v|c v¦v c ¦v a|c u¦u m
C|C	" "	¦v c|c v c ¦u n|d e r
C|CC	" "	¦v c|c c v ¦e n|t r y
CC|CC	" "	¦c c v c c|c c v c c ¦g r a n d|s t a n d

1st syllable boundary.

No syllable boundary between initial/final consonants.

*(Most vowel sequences are not divisible, by rule.)

each one, according to stress level. (Consonantal quality is also affected by stress, but not as profoundly as vocalic quality.)

But in order to assign stress, syllabification must first be done. Slide E outlines the method used.

At present we are assuming that printed letters can be recognized, so we start by simulating the reading machine's functions in stress and phonetic assignment. This generates, by rule, the phonetic string that serves as the input to the computer and synthesizer.

In the simulation process, we consult a list (as it might exist in computer memory) which contains the stress and phonetic information for about 100 affixes, that match spelled letter sequences. We also store about 500 common words, because they differ, in spelling or stress pattern, from the vocabulary at large. For example, articles and short prepositions are stored with low stress. Words such as "talk" and "child" which are not pronounced as the spelling suggests, are stored; and words like "politics" which has unusual stress considering its suffix, are also stored. Slide F shows the steps we use in converting print to phonetics:

I. Test entire word for match in storage.

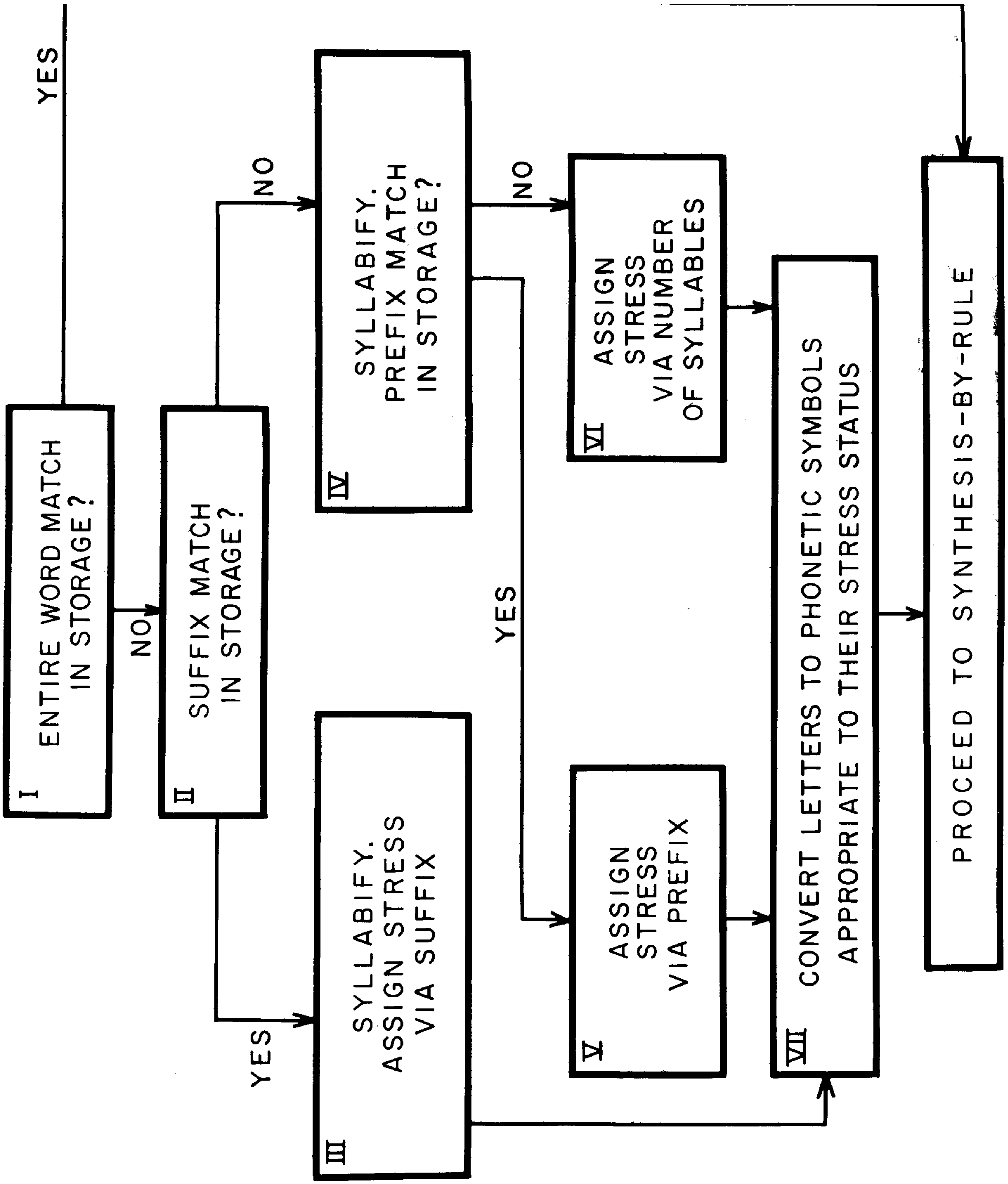
II. If no word match, search for a suffix match.

III. If suffix matched, syllabify and assign stress on basis of suffix.

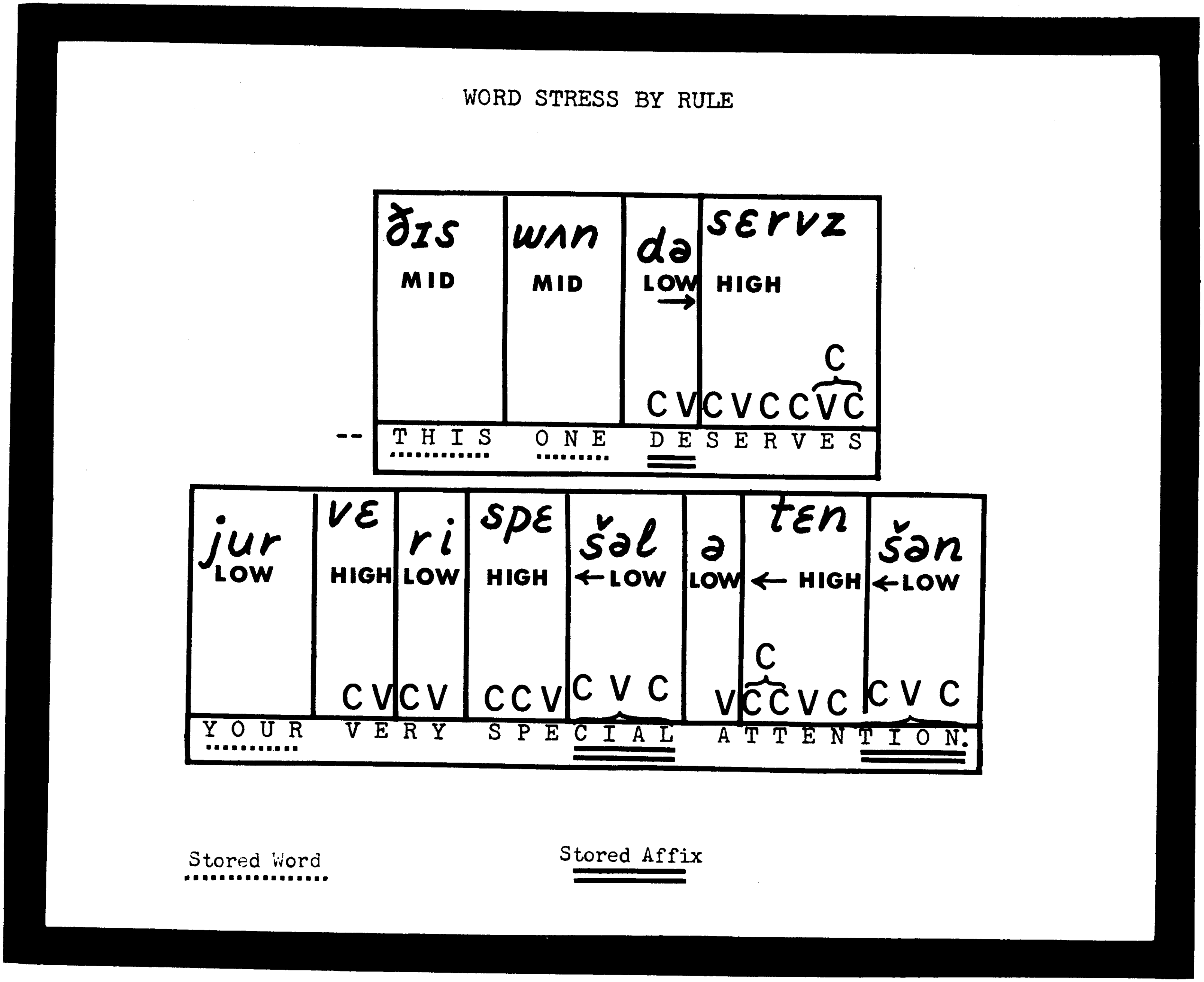
IV. If no suffix match, syllabify and search for prefix match.

V. If prefix match, assign stress on basis of prefix.

VI. If no prefix match, assign stress on basis of number of syllables in word.


I ENTIRE WORD MATCH IN STORAGE ?
YES
NO
II SUFFIX MATCH IN STORAGE ?
YES
NO
III SYLLABIFY. ASSIGN STRESS VIA SUFFIX
IV SYLLABIFY. PREFIX MATCH IN STORAGE ?
YES
NO
V ASSIGN STRESS VIA PREFIX
VI ASSIGN STRESS VIA NUMBER OF SYLLABLES
VII CONVERT LETTERS TO PHONETIC SYMBOLS APPROPRIATE TO THEIR STRESS STATUS
PROCEED TO SYNTHESIS-BY-RULE

SLIDE F


WORD STRESS BY RULE
ðɪs wʌn də sɛrvz
MID MID LOW HIGH
C
CV CVCCVC
-- THIS ONE DESERVES
jur vɛ ri spɛ šəl ə tɛn šən
LOW HIGH LOW HIGH ← LOW LOW ← HIGH ← LOW
C
CV CV CCV C V C V CCVC C V C
YOUR VERY SPECIAL ATTENTION.
Stored Word
Stored Affix

SLIDE G

VII. Convert letters to phonetic symbols appropriate to their stress status.

Slide G shows the stress and phonetics assigned by rule to a sample sentence. The first tape sample you will hear was produced from this phonetic string, used as the input to the computer and synthesizer. Ignatius Mattingly's program for the synthesis of American English by rule -- including intonation -- did all the rest.

References

Bertrand, Georges. L'accentuation lexicale des polysyllabes anglais: Finales à accentuation fixe. In In Honor of Daniel Jones, D. Abercrombie et al., Eds. (Longmans, Green and Co., Ltd., London, 1964) 223-241.

Delattre, Pierre. Comparing the prosodic features of English, German, Spanish and French. Internat. Rev. Appl. Ling. in Lang. Tchg. I, 3-4 (1963).

Waldo, George. The significance of accentuation in English words. In Proceedings of the Ninth International Congress of Linguists, Cambridge, Mass., 27-31 August 1962; H.G. Lunt, Ed. (Mouton and Co.) 204-210.