The Machine Conversion of Print to Speech: Two Papers

JANE H. GAITENBY

Miss Gaitenby is on the staff of Haskins Laboratories, New York City.

The following two articles deal with some of the linguistic aspects of devising an audible verbal output for a print-reading machine for the blind. The machine, which is now under development, will have a direct input of printed material—a textbook page, for example—and will produce an output of audible spoken sentences. The prototype machine is planned to be suitable for a central reading facility, such as a library, and is not designed for portability or individual ownership. It is, however, intended to be accessible—in the future—to individual users by means of a telephone line to the library.

☐ The first article, "Problems in Machine Conversion of Print to Speech," describes experiments in compiling sentences from pre-recorded single spoken words. This is one type of possible output for a reading machine. The second article, "Rules for Word Stress Analysis for the Conversion of Print to Synthetic Speech," deals with software for another type of spoken output: speech made entirely by machine.

These articles were originally presented as lectures, accompanied by taped examples of the output of the reading machines. Interested readers can obtain copies of these tapes by writing to the *New Outlook for the Blind*, 15 West 16th Street, New York, N.Y. 10011. For technical reasons, there may be a few weeks' delay in filling these requests. There will be a charge of $1.50 to cover cost of materials and mailing.

A key to the demonstration tapes appears at the end of the second article.

Human and Machine-Made Speech

I. Problems in the Machine Conversion of Print to "Speech"

This paper was presented in March, 1966, at the 11th Annual National Conference on Linguistics in New York City. It was first published in the Spring 1966 issue of the Bulletin of Prosthetics Research.

Applied research that is linguistic in character is required by a variety of enterprises and institutions these days. The research to be reported here has been made under contract to the Prosthetic and Sensory Aids Service of the Veterans Administration. The Veterans Administration has supported a number of research projects for the purpose of developing reading machines for blind persons; among them have been studies in electronic instrumentation, in psychology, in linguistics, and in cross-disciplinary fields. Each of these investigations has been concerned with the problem of converting the printed word to a tactile or auditory output. Part of the general research problem is optical in nature (because print must be recognized electronically), part deals with printed-symbol-to-sensory-symbol correspondence, and another part deals with human perception of units of sounds. A specific problem area is the analysis of the structure of spoken English, in view of the fact that the printed word is speech at one symbolic remove. Since blind people in general are obliged to approach all handwritten or printed or illustrated material either through braille or through the intercession of a human reader, a reading machine of some kind is an obvious need.

The Machine Should Talk

☐ Leaving the consideration of cost aside, we can assume that the best reading machine will be one that converts printed text to an output that is as much like real speech as possible. In short, the machine should talk. This ideal machine should produce completely natural sentences, and to do this it should have the ability to vary intonations and pauses appropriately for specific texts. Ideally, it should be replete with variable voice quality, such as one shade and timing of voice for business letters and another one for romantic novels—or letters. But such a perfect machine would require a gigantic storage capacity, starting with tens of thousands of words. It would also require a remarkable program to manipulate its memory—to account for all manner of related nuances: grammatical, semantic, and intonational—in order to duplicate the associational memory and variable voice of the human being. Bear this ideal machine in mind—and the fact that it *is* an *ideal* machine.

Sentences from pre-recorded words

Speech has been synthesized at Haskins Laboratories and elsewhere by rules applying to very small speech units, generally, of phonemic or syllabic size. A logical developmental step was to experiment with sentence production from work units that had been pre-recorded by a human speaker. A device called the interim word-reading machine was built for the Veterans Administration at Haskins to test the feasibility of generating sentences from single spoken words. Although a successor to that machine is already under way, many of the problems encountered in the course of outfitting the first interim device with a vocabulary, and other equipment, remain; so I will confine my remarks to the machine that has already served much of its purpose, but is soon to become ancestral.

The well-named *interim* word-reading machine at Haskins deals only with the word storage, retrieval, and output side of the reading machine problem, and omits the optical scanning operation (which is not our contractual obligation). My concern has been to get words recorded for the machine's storage, words that can be played out one after another, to (presumably) sound like sentences. The print-scanning function has been bypassed in the machine by simulation. The contents of a printed text are typed on a Flexowriter. The punched-tape resulting is simply a letter-to-digit conversion, and this is the form of information put into the machine. This is the input.

Vocabulary Storage

☐ Now about the vocabulary storage in the machine. Each separate spoken word that has been previously recorded is stored on magnetic tape. Stored along with each word on the tape is its digital spelling, or code. When the code for a given word is sensed on the *punched* tape by the machine, that word is searched out on the stored spoken vocabulary *magnetic* tape. The word is matched and played back if it is stored, and is recorded on another tape at the same time, where it is added to the rest of the words in the order commanded by the original text. (If the word is not stored, it must, unfortunately, be spelled out, letter by letter.) When the contents of the entire text have been accumulated, this new tape, consisting of spoken words, is played back to the listener. This is the output of the word-reading machine, and the entire process is, in fact, a conversion of print to sound.

The best feature of the interim reading machine output is its approximation of natural intonation, particularly in prepositional phrases. Since phrases of this type occur in texts far more often than any other syntactic structure, the overall acceptability is high.

Spelled words hard to understand

The most unnatural aspect of the synthetic sentences is the presently unavoidable number of spelled words, occasioned by the restriction on the size of the stored vocabulary. When, in the midst of the verbal output, the listener suddenly hears a spelled word, he is somewhat bewildered and his comprehension of the sentence suffers. Spelling, since it is not typical of either conversation or of reading represents a total shift from the medium of speech to the medium of writing. Spelling is therefore destructive of the intelligibility of the whole text, despite the fact that it is necessary in a mere five percent of the words of a normal text. In addition, the unrecorded words which must be spelled in the output are the most infrequent (i.e., least expected, least guessable) words. And they are—on the average—exceptionally long, or peculiarly spelled, "difficult" words. Furthermore, each letter of a spelled word is a whole syllable, and thus the output word rate is severely slowed down by spelled words. Another handicap in spelled words is that the letters, as spoken, lack natural intonation and spacing, which a good spontaneous reader would provide if he were obliged to spell. Each of the 26 letters had to be recorded in only one intonational form, just as each of the 7,200 words was recorded in a single spoken version. Because there are even fewer restrictions on the place of letter occurrence in a word than there are on the place of word occurrence in a phrase or sentence, advance prescriptions for specific letter intonation were not attempted. The letters were merely recorded in groups by rhyming syllables, e.g., "A,J,K . . .," at a sustained single level, in the hope that in sentences they would rapidly be perceived as letters, distinct from the words spoken as units in their immediate environment. A very brief pure tone signal also precedes and follows each spelled word in the program for the output—to signal the abrupt shift from whole words to letters on the auditory track—and this seems helpful.

As mentioned just above, the total recorded vocabulary consisted of 7,200 units: words, letters, numerals, a few suffixes, and reflexes for punctuation—both spoken and silent. The original words were spoken by John T. Wadsworth, in a long series of short recording sessions which were made over the course of a year.

Words mounted on cards

After the words were recorded, the original tape was copied and edited. Then each word was separately mounted on a card. The test sentences were generated by playing the word cards in sequence through a machine (Language Master) that plays tapes from cards instead of reels, and then each word was recorded in direct sequence on another machine. The sentences *as such* are therefore synthetic, since they were never spoken as sentences by the original speaker. The output demonstrates by negative evidence that real speech is produced not word by word, but in continuous groups of words that are compatible in respect to tempo, loudness, and melody.

Each Word Stored Only One Way

☐ Every problem in applied research has constraints. For this machine, one—and only one—intonational version of each vocabulary word could be stored. This is a severe restriction, because in normal speech a given word may occur in many very, *very* different prosodic forms. Also, just one pronunciation was allowed per printed form, but homographs are a minor problem, compared to the fact that each word had to be stored in a single, frozen, stress and intonation form. It is clear that the one version recorded should be a highly probable spoken form for a highly probable type of occurrence in printed form.

We chose the vocabulary to be stored from the Dewey and the Thorndike and Lorge lists of the most frequently printed English words. There were no published data on probable spoken forms of the vocabulary words, although there were helpful reports in the literature on measurable correlates of stress and intonation by such investigators as Bolinger, Fry, and Denes.

Stress relationships in polysyllabic words

Dr. F. S. Cooper of the Haskins Laboratory gave us a take-off push in the right general direction and some fine instruments to work with. Having made exploratory acoustic and perceptual texts of real and synthetic speech, we approached the machine's speech problem with the hypothesis that very frequent phrases and polysyllabic words were structurally similar in a way that might be useful to us. That is, the syllables of a polysyllabic word have a persistent stress relationship, even in varied intonational environments, and the syllables of the most frequent phrases also tended to be regular in their stress relationships. We gradually learned something about the prosodic components of stress in words or phrases—intensity, frequency, and duration—and decided to "program" our human speaker, if possible, to make the word recordings using prescribed stresses and intonations. Now the question was, what stress to prescribe for what. We turned to the study of the probability of word occurrence in printed and spoken form.

The most frequent type of phrase in English texts is the prepositional. Articles, prepositions, and a connective are by far the most common English words. Most highly frequent words are monosyllabic. The stress of the most frequent words is usually very low. Most phrases begin with a preposition, as mentioned before, and most phrases and sentences end with a noun. Nouns carry much information and are generally prominent in the speech chain—and so on with the other grammatical classes.

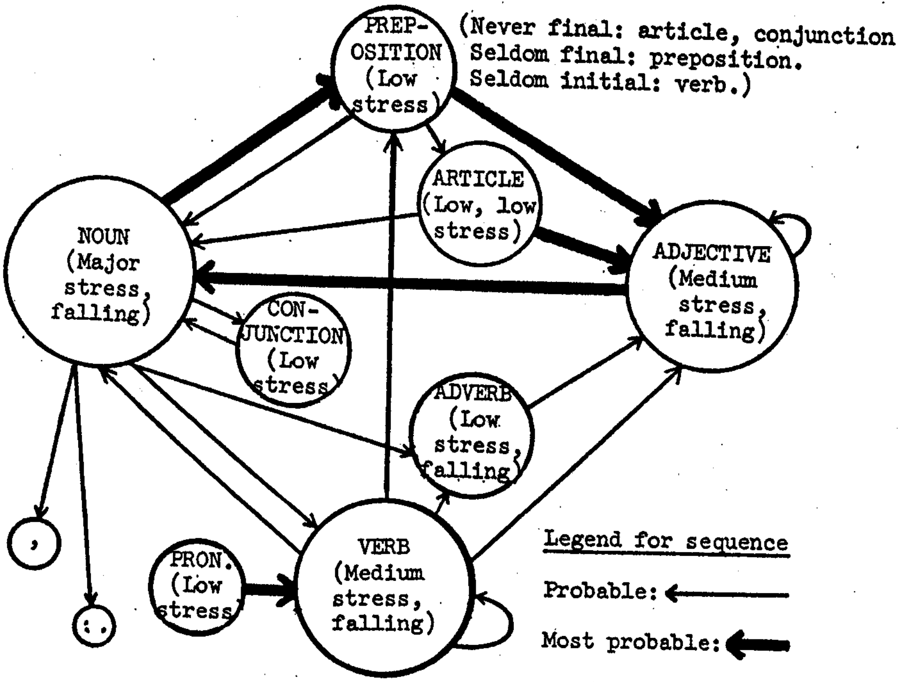


FIGURE 1
Hypothesized Most Probable Grammatical and Intonational Sequences in English Texts

We also examined so-called "intonation" and were soon convinced that it is acoustically reflected in durational shifts as well as in frequency and intensity changes. Otherwise, we knew very little except that syllables lengthen immediately before a pause, that pitch and intensity peaks usually start high and tend to decline toward the end of an utterance, and that pauses for punctuation are variable in length and have structural significance (Some of these aspects of intonation almost certainly have a physiological base).

Words Recorded by Grammatical Function

☐ Facts or observations of this sort, along with counts of form class sequence taken from texts in daily papers, books, and periodicals, indicated that the spoken lexicon should be recorded on the basis of a word's grammatical function. Grammatical function is correlated to some extent with word stress category and a trained speaker could, with effort, produce words at a prescribed relative pitch, length, and loudness. An underlying assumption was that the stress prescriptions themselves would be valid, and that they would be consistently reproducible by a human speaker on demand, and over a large span of time.

Figure 1 is a diagram of probable grammatical sequences in printed texts. The size of each circle indicates the projected relative prominence of a word as a given part of speech, used in writing the stress prescriptions. Although nearly all the form classes in English can be—and often are—preceded or followed by almost any of the form classes, the arrowed lines shown between any *two* classes stand for statistical likelihood of sequence.

In order to use the probability rationale for the manner in which word classes were to be spoken, we had to classify each of the 7,200 words as a member of a particular grammatical class, before it could be suitably recorded. This was a stumbling block.

Words Listed by Probable Function

☐ Table 1 suggests the problem by showing a breakdown of the thousand most frequent words by potential membership in a grammatical class. About half of the words can function in more than one role, and that role can be determined only by context. We therefore classified the multifunctional words by their most *probable* function, through educated guesswork and by intuition as native speakers. Impossible-to-classify words were put into the most neutral stress group, along with main verbs, whose stress in a sentence seems to be unpredictable. The prescriptions were written, and the recordings were made.

The recording speaker followed the directions explicitly about 99 percent of the time, and the texts used have actually been a random test of the stored 7,200 word vocabulary. The questions to think about when listening to the tapes are: To what degree is normal intonation approximated in these tapes, in what kinds of cases is it least normal, and what *is* normal intonation?

TABLE 1
Grammatical Functions of the Most Frequent Words

A) Words that have a *single* function,		
Nouns	196	
Verbs	140	
Adjectives	81	
Adverbs	48	
Prepositions	17	
Connectives	8	
Other	1	
	491	
B) Words that have *two* potential functions,		
Noun or Verb	215	e.g., "being"
Noun or Adjective	120	"three"
Verb or Adjective	50	"open"
Adjective or Adverb	26	"better"
Adverb or Preposition	18	"to"
Adverb or Connective	8	"since"
Connective or Other	6	"though"
Adverb or Other	4	"really"
	447	
C) Words that have *three* potential functions		
Noun, Verb, Adjective	46	e.g., "present"
Noun, Adjective, Adverb	7	"first"
Noun, Adjective, Connective	3	"either"
Verb, Adjective, Preposition	3	"near"
	59	
D) Words that have *four* potential functions,		
Noun, Verb, Adjective, Adverb	6	"last, left, set, front, further back"

From Printed Word Direct to Audible Syllables

☐ Studies are now underway to formulate a program that will circumvent the spelling problem by converting printed words directly to audible, recognizable syllables. It is well known, however, that there are few one-to-one spelling-to-sound conversions in most of the letters and syllables of English (especially because of stress problems). Thus, this experimental method presents problems of its own.

The second generation word-reading machine now being developed, sired by the interim machine and a computer facility, will store words as control signals rather than as waveforms as at present. In that machine, intonation can be manipulated and programmed by synthesis, and certain of the present output irregularities in stress, timing, and inflection, will be reduced—if not completely eliminated. The word rate can be speeded up as the phrasing is improved, and this is a requisite for blind listeners. At the same time our understanding of real speech phenomena will increase.

II. Rules for Word Stress Analysis for Conversion of Print to Synthetic Speech

This paper was presented at the 74th Meeting of the Acoustical Society of America, held in Miami Beach in November, 1967.

Learning to read often takes humans several years. In contrast, a machine that reads print aloud must be an accomplished reader right away, although it has had no language, spelling, or writing practice, or general skill development in contextual recognition. The human reader listens, associates, and learns to talk long before he learns to read, and learning to read is a gradual thing, just as learning to speak was, before it. The human being who is being taught to read must learn the following subsidiary skills:

1. To recognize printed letters, both upper and lower case; and to be able to spell, pronounce, (and have connotations for) English words, by knowing both their regularities and irregularities in specific instances.
2. To apply word stress correctly.
3. To supply normal intonation to printed sentences, in which the printed guides to spoken inflection are largely cryptic, or non-existent—if meaning is considered apart from the printed signal. The chief cues to intonation in printed texts are punctuation and word order. The other signal, semantic content, is provided by the human reader's personal experience in speaking and hearing English. (The human consults his auditory memory to provide proper intonation to a text.) A huge associational memory for spoken language is a very special advantage that all non-deaf humans possess, unlike the comparatively puny memory (storage) that can be built into computers at the present time.

The Reading Machine Doesn't Understand

☐ The human reader is expected to comprehend what he reads, of course; but reading machines do not understand a syllable of what they must convert to sound. A computer program is required to provide the reading machine with quick strategies to make decisions about word, phrase and sentence pronunciation. Therefore the program must contain procedures for reading in a meaningful-sounding way; i.e., it must be based on the linguistic realities of spoken English.

To begin with, the machine must match or identify the printed letters. (This is done by optical character recognition—a step that will be omitted here because it has not been a part of our contractual obligation). After letter recognition has been achieved (or simulated, as in our model) the machine may use sequential reading procedures, each roughly corresponding to the mastery of a human step in learning to read. The reading steps used by the machine are stored data and operations pertaining to the following subjects:

1. The most probable pronunciation for each letter and for various frequent letter sequences;
2. The most probable stress of syllables as words, parts of words, or parts of phrases;
3. The most probable intonation for sequences of words which are terminated by specific punctuation symbols (and perhaps, as framed by particular word classes).

The voice must sound human

A reading machine that produces a synthetically-made spoken output must, in addition, speak in a voice that is not un-human, so that the speech will be pleasant to listen to, as well as normally intelligible.

As has been said earlier, the machine's speech must also *sound* as if the machine understands what it is giving voice to, although its performance is entirely automatic and without meaningful reference to either content or context. The (blind) listener must be enabled to derive the full sense of utterances from the machine's delivery, even though the speech is synthetic and the machine's "interpretation" is, literally, literal.

The task of reading by machine is comparable to the human task of reading nonsense words aloud. The machine's word pronounciation must be based on the spelling, and its sentence intonation based entirely on punctuation. Nevertheless, the machine is intended to produce an intelligible spoken text for human listeners—and for blind listeners in particular.

FIGURE 1
Print Reading by Machine

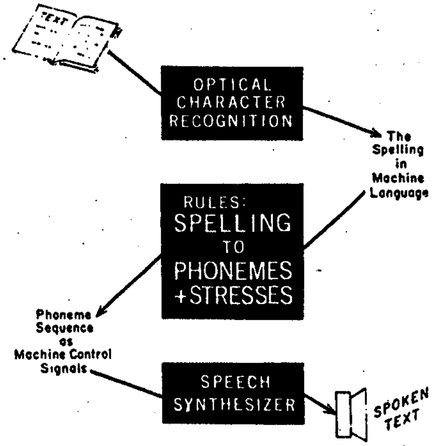


Figure 1 shows the reading process as it is to be done by one type of machine (the one that will concern us here). The three black boxes represent its main operations. An optical character recognizer transforms print to machine language; then (as seen in the center of the diagram) the spelled word becomes the phonetic word, *by rule*. (The rules for word stress analysis apply to this step.) Finally, the phonetic symbols—called "phonemes" in Figure 1—become commands for synthesizing speech. This last step involves speech synthesis by rule, which is a rapidly advancing art and science as has been demonstrated by Drs. Ignatius Mattingly and Mark Haggard, and others. Mattingly and Haggard have devised programs for the synthesis of American English and British English, using the Haskins Laboratories computer and synthesizer. The form of computer input they have used is 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 (sounds). We are not alone in this attempt, but there are problems enough in English spelling to go around.

A nonsense-word test

As a start, we set up a test of nonsense words for humans to read, so that we might observe their approach to a reading problem which would be equally difficult for man and machine. The general purpose was to determine the *kinds* of pronunciation cues—and the *order of cues*—employed by humans when reading meaningless syllables. The nonsense items resembled English words: normal letter sequences were used, and many of the words contained familiar affixes. Figure 2 shows samples of the words used.

The specific purpose of the test was to see how people stress unknown words—and in particular, to pinpoint the role of affixes in the assignment of word stress. (Stress, sometimes called "prominence," is a speech property that can be perceived in any spoken syllables. There are numerous degrees of stress, and all of them are relative. In describing speech, linguists recognize at least three—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 unusually low) in voice pitch, or longer and louder than its neighbors. In addition, it may be set off by silence from its neighbors.

Returning briefly to the nonsense-word test: A total of 150 words, including those shown in Figure 2, were presented in four test sessions to each of 12 adult readers. The assumptions to be tested 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. Figure 3 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 a two–or three–syllable word.

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

FIGURE 2
Sample Nonsense Words in Tests

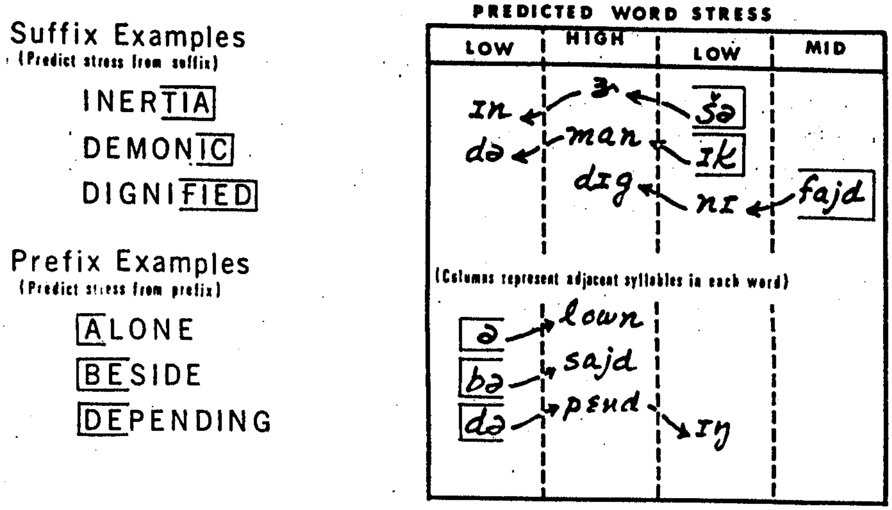


FIGURE 3
"Stable" Affixes and Stress Prediction

3. Successive syllables in a word tend to be pronounced with alternating high and low stresses—unless the root, or the affix region, is compound.

4. Stress alternates backward from a stable suffix (look at Figure 3 again), or forward from a stable prefix. When both a suffix and a prefix are present and stable, the suffix determines the word stress.

5. The stress pattern attributed to a word *limits* the phonetic quality given to its syllables, the *vowels* being most noticeably affected; e.g., "*in*valid" with inverted alternating stress becomes "in*va*lid."

Assumptions Confirmed

☐ All of these assumptions were substantially confirmed. The assumptions, which now serve as working rules, are based on the enormous frequency of stable affixes in actual texts, and on the tendency of stress to alternate in the syllables of a word. (High—Low—Mid-High—Low is the general pattern of alternation.) Further, the habitual stress level of a stable affix allows us to extrapolate the stress of the rest of the word.

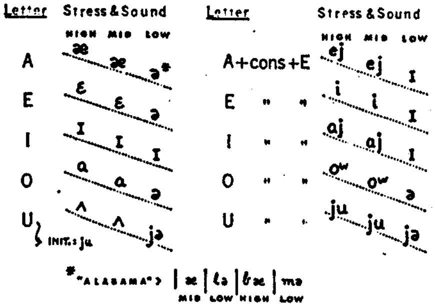
If we combine stress probability with the observation that vowel color reflects the level of syllable stress quite dependably, we have a foothold on the concept of print-to-speech conversion. Figure 4 shows the single vowel letters, along with three phonetic states probable for each one, according to 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. Figure 5 outlines the method used.

At present we assume that character recognition has already taken place as the first step in the machine's reading process, so we start in by simulating the reading machine's functions in stress and phonetic assignment. This involves rules for generating 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

FIGURE 4
Sound Correspondences for Single Vowels, by Syllable Stress Type



Letter	Stress & Sound HIGH	MID	LOW	Letter	Stress & Sound HIGH	MID	LOW
A	æ	æ	ə*	A+cons+E	ej	ej	ɪ
E	ɛ	ɛ	ə	E " "	i	i	ɪ
I	ɪ	ɪ	ɪ	I " "	aj	aj	ɪ
O	ɑ	ɑ	ə	O " "	oʷ	oʷ	ə
U } INIT.: ju	ʌ	ʌ	jə	U " "	ju	ju	jə

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

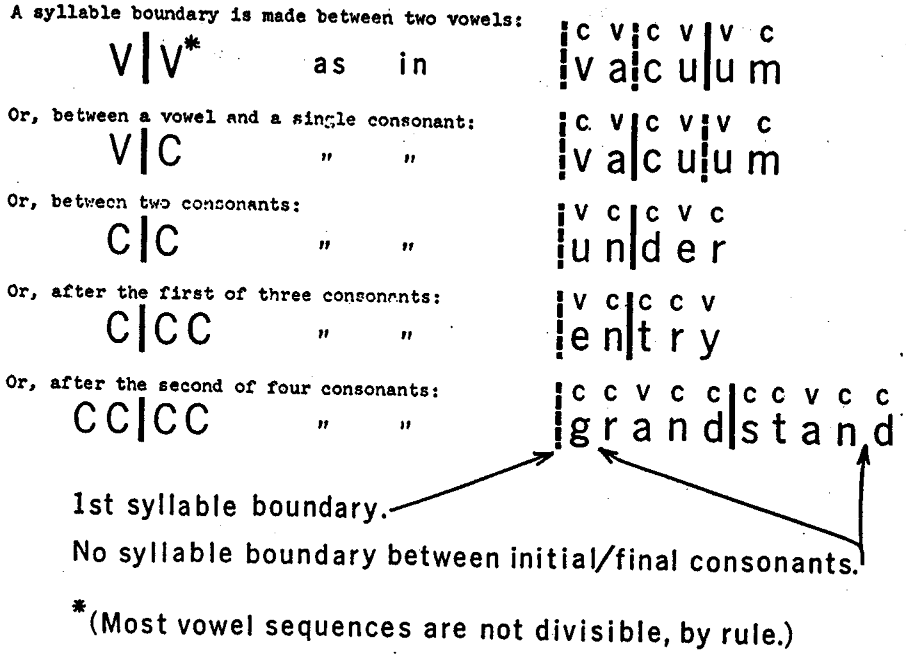


FIGURE 5
Syllabification Rules

100 affixes (that match spelled-letter sequences). We also consult a list of (i.e. 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 assigned to them (since they are customarily unstressed in speech) and words, such as "talk" and "child," which are not pronounced as the mere spelling suggests, are stored. Words of high-usage frequency such as "politics"—which has an unusual stress pattern in view of its suffix, are also stored. (Usually the ic(s) suffix denotes full stress on the preceding syllable—e.g., "platonic." In "politics," the syllable preceding the suffix is unstressed.)

Stress Can Be High, Mid-High or Low

☐ Another part of the storage consists of phonetic specification for the most probable type of sound corresponding to each alphabetic letter (and each common short letter sequence, e.g., "ea," "sh") for High, Mid-High, and Low Stress states.

Figure 6 shows the steps followed in converting print to phonetics, here elaborated in the text:

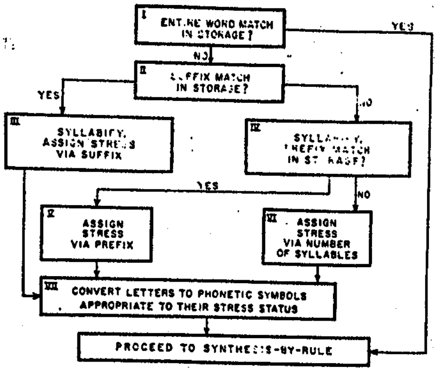
1. Test entire word for match in storage.
2. If the word is not matched, search for a suffix match.
3. If the suffix is matched in the storage, syllabify the word and assign stress to the syllables of the word on the basis of the suffix instructions. (For example, if the suffix is stored as Low in stress and signals High Stress in the preceding syllable as the suffix "-ion" does, the preceding syllable is assigned High Stress, and the syllable preceding *it* receives Low Stress, etc.)
4. If the suffix is not matched, syllabify the word and search for a prefix that matches the first syllable.

5. If the prefix is matched in storage, assign stress on the basis of the prefix instructions.

6. If there is no prefix match, assign stress on the basis of the number of the syllables in the word. (Monosyllables receive Mid-High Stress, disyllables receive High—Low, trisyllables receive High—Low—Mid-High.)

7. Convert the letters of the word to phonetic specifications appropriate to the stress status of the syllables in which they appear.

FIGURE 6
Stress in Converting Print to Phonetics

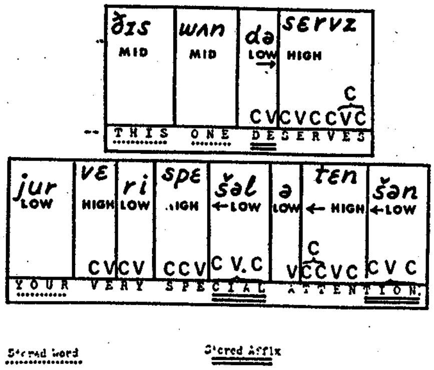


In Figure 7 are shown the stress and phonetic symbols assigned by rule to a sample printed sentence. (At this point in the oral presentation of this paper, a tape recording of several sentences was played to let the audience hear how the actual print-to-synthetic speech conversion sounded after the rules had been processed through a computer and speech-synthesizer hook-up, using Mattingly's synthesis algorithm for American English, which included a formula for supplying intonation to the utterances.)

If the reader is familiar with phonetic symbols, Figure 7 will be illustrative. If not, a word of explanation of the figure and the tape it represents is in order:

Listeners who heard the tape demonstration found that the sounds produced for the symbols (in Figure 7) were close to what humans say aloud for the printed words in the text, except that the word "deserves" was pronounced as "deSAIRVZ" because there is no simple rule for midword conversion of the letter s to the sound z, or for "er" conversions to the proper sound for this particular word (rules for vowel + r sequences in stressed syllables present many problems). Listeners also noted that the intonation was a little extreme for natural American speech, and that the voice quality of the machine did have a mildly mechanical air—although the total sentence was fully intelligible.

FIGURE 7
Word Stress by Rule



Rules for word stress analysis and assignment, resembling those presented in this report (or augmented by syntactic information), have implications for teaching humans to read, in addition to their clear usefulness in programing machines to read. At present the rules produce some results that should never be produced by good human readers, but are often produced by poor human readers. The occasional negative results in phonetics, stress assignment and intonation can be surprisingly instructive to both the linguist and the teacher of reading.

Overall, the relative simplicity and actual utility of rules for correctly stressing the majority of English words has been demonstrated. In controlled experiments (such as the one cited) a simulated storage of stress rules has been applied to Mattingly's synthesis program, and a speech synthesizer has produced audible and intelligible sentences from (simulated) inputs of printed texts.

A Key to the Demonstration Tapes

☐ This key is presented for readers who wish to order copies of the demonstration tapes, and for subscribers to the recorded edition of the *Outlook*.

The first four sections of the tape illustrate the technique discussed in the first article—pre-recorded single spoken words and letters, using the human voice. The fifth section illustrates the technique discussed in the second article —speech sounds generated completely by machine.

Section 1: "This is speech. . . ." Pre-recorded single words.

Section 2: "Every city has its. . . ." Pre-recorded single words, with letters used to spell words not stored in the machine. Played 20 percent faster than in section 1.

Section 3: "The north wind. . . ." Pre-recorded single words, no words spelled out. Played 20 percent faster than in section 1.

Section 4: A personal letter. Output includes punctuation marks and numbers. Same speed as in section 1.

Section 5: Speech made entirely by machine, no human-voice input. Near the end of the section, the machine pronounces the French phrase, *"c'est tout fini."*

Bibliography

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

Delattre, Pierre: "Comparing the Prosodic Features of English, German, Spanish and French." *International Review of Applied Linguistics in Language Teaching,* 1, 1963, 3-4.

Waldo, George: "The Significance of Accentuation in English Words" in H. G. Lunt (ed.), *Proceedings of the Ninth International Congress of Linguists* (Cambridge, Mass., August 27-31, 1962), 204-210. Mouton and Co.

Reprinted by permission of
THE NEW OUTLOOK FOR THE BLIND