REPRINTED FROM Volume I—*Proceedings of the International Congress on Technology and Blindness*

PSYCHOLOGICAL CONSIDERATIONS IN THE DESIGN OF AUDITORY DISPLAYS FOR READING MACHINES

MICHAEL STUDDERT-KENNEDY* *and*
ALVIN M. LIBERMAN**
Haskins Laboratories, New York, New York

Our concern in this paper is with the choice of an auditory output for a reading machine. We shall consider the choice primarily from the point of view of the blind person who is to use it. This may mean that we end by throwing the burden on the engineer, but that is what engineers are for, after all, and it is not unreasonable that the blind user who must listen to the machine should have a say in what it sounds like. So it is on the blind person's needs and human limitations that we shall dwell.

First, his needs. They are simple. He needs a device that will permit him to read at a rate, if not as high at least of the same order as that of the sighted reader. A reasonable aim would perhaps be the rate of normal speech, that is, about 150 to 200 words per minute. Such a rate is necessary for two reasons. First, slow reading is irksome, especially when, as is often the case with newspapers and magazines, the ratio of content to text is deliberately kept low to facilitate fast reading, easy comprehension, and mass circulation. Second, slow reading is inefficient; immediate memory is short and the slow reader may forget the beginning of the paragraph or even the sentence before he reaches the end. The fast reader takes in the sweep even if he occasionally misses the detail; the slow reader may take in neither sweep nor detail. In the interest, then, of both pleasure and efficiency the blind reader may reasonably ask for a device that will read to him at least at a normal speaking rate.

We are not, of course, denying the utility of less ambitious devices. A cheap, portable machine, even of limited performance, has obvious value for the personal use of many blind persons, and we have no wish to be-

* Also at Barnard College, Columbia University, New York
** Also at the University of Connecticut, Storrs, Connecticut

little research efforts directed toward the development and improvement of such devices. Here, however, we have chosen to confront the larger challenge of a high performance reading machine, the library installation, that will give the blind reader access to the world of books from which he is presently shut out.

Turning now to the larger matter of the listener's limitations, we find it illuminating to compare spoken with written language. Speech, the primary medium, is a continuous flow of sound arrayed in time; written language is its discontinuous visual representation arrayed in space. This discontinuity or segmentation of written language is the crux of our problem. Alphabets were surely good solutions to the problem of segmenting the acoustic stream for symbolic presentation in nonacoustic form. But from our immediate point of view the solution was all too good, since we are now faced with the reverse problem of how to put the pieces together again.

How does the sighted reader do this? First he scans the spatial array of print and in so doing transforms it back into a temporal array. But he does not do this continuously; though the subjective impression may be of his eyes steadily sweeping the print, in fact they are moving in small discrete leaps, called saccadic movements, each lasting some 10 to 40 milliseconds. Between movements his eyes fixate the print and it is only during fixation that retinal stimulation is effective. The number of fixations per line varies with the reader and with the material being read. A good college freshman studying a text of average difficulty will fixate four or five times a line. Each fixation will last around 200 to 250 milliseconds and total fixation time will be 90 to 95 percent of total reading time (19). During each fixation the reader "takes in" several words. Since the fovea, the retinal region of highest acuity, covers a visual angle of approximately 2 degrees, only one or two letters are clearly in focus; the letters peripheral to the fovea are somewhat blurred. The reader takes another peripheral look at the blurred letters during his next fixation and thus achieves a visual substitute for the continuity of speech. Strictly, of course, he is still segmenting the flow. But by grouping lines into letters and letters into syllables, words, and phrases during fixation and by cutting the dead time of saccadic movement between fixations to a minimum, he effectively achieves a continuous intake at a rate even faster than that of spoken language.

We may remark that exactly the same sequence of rapid eye movement interspersed with relatively long fixations is followed when we look around us and group the chaos of the external world into people, trees, houses, and objects. We do not, of course, see every detail any more than the

reader, but by long practice we have learned to focus on the essential cues—brightness contrast, continuity of line or surface, pattern repetition—and so broadly organize the visual world "at a glance," as we somewhat inaccurately say. It is this spatial span of vision that the automatic scanning guidance device is unable to recover from the free field. Perhaps a reading machine can do a better job for the printed text.

We have stressed the organizing power of vision, and we have implied that for a high performance reading machine we must provide some auditory counterpart of this power. The question now arises as to where the organization is to take place, in the machine or in the listener? To be more exact, how much organization should take place in the machine and how much in the listener?

If we place all the burden of organization on the listener we are choosing what has been called a direct translation, nonintegrating type of machine; the Optophone is a famous example. Such a machine presents the listener with a series of sounds that "are generated from, and vary in accordance with, the continuously changing contours of the print" (2). Certainly this is a faithful translation of the text, but how unlikely it is to yield high reading rates one can judge by imagining oneself visually reading a text through a slit so narrow as to offer only a fragment of a letter at a time. One's level of organization would hardly develop beyond the letter. And we surely want our translation to catch more than the letter of the text.

There is a more important reason why we cannot expect optimal reading rates from such a device, and this is the fact that the ear has relatively low resolving power in time. A series of clicks or other brief sounds are heard as discrete only so long as they are not repeated more than about 20 times a second. Above that rate they merge and are heard first as a buzz and then as a steady tone of rising pitch. Clicks merge into a buzz even when they differ from each other in frequency composition, intensity, and duration. In International Morse Code, for example, each letter is coded into a distinctive series of long and short wave trains of a 1000 cps tone with an average of three wave trains per letter. Since the average English word contains 5 letters (or 15 Morse Code elements), a rate of 1 word per second or 60 words per minute is close to buzz threshold and the presumptive upper limit for Morse Code reception. In fact, operating rates are considerably lower than this: commercial radio stations send and receive at no more than 30 to 40 words a minute, federally licensed radio amateurs at 13 words a minute (2).

An upper limit of roughly the same rate may be expected from such

devices as the Optophone. In fact, the most proficient user of the Optophone, Miss Mary Jameson, gave public demonstrations of reading at reported rates of 60 words a minute(2). Other users achieved much lower rates.

The fact that most users of such codes achieve rates so far below the theoretical upper limit is, incidentally, of some interest. Both International Morse Code and the Optophone use unidimensional codes; that is to say, the letter symbols differ from each other along a single dimension—duration pattern for Morse Code, frequency for the Optophone. Pollack (16, 17) and others (5, 11, 14) have shown that listeners identify more accurately stimuli that differ along several dimensions, and it may be that some improvement in rate could be achieved by suitable complication of the signal. Of course, increased accuracy of identification may not yield increased speed of identification, but Eriksen (5) has found an increased rate of identification for complex visual stimuli under certain conditions, and it is possible that a similar gain might be achieved with auditory stimuli.

Be that as it may, no amount of stimulus complexity can overreach the ear. There is a physiological upper limit to the rate at which a listener can discriminate between successive discrete auditory stimuli, and if he cannot discriminate between them he clearly cannot organize them. Our first conclusion, then, is that we cannot place all the burden of organization on the listener; the machine must do something, too. The question is, how much should it do?

Let us recall the sighted reader. His first level of organization is the letter. If we assign this task to the machine we are choosing what has been called a recognition, letter-reading type of machine (2). Such a machine identifies each letter discretely and yields a corresponding sound. The sound need not resemble the letter, but the task of the listener is simpler if he does not have to learn a new code, and letters are the obvious choice. To attain reasonable speeds the letters or their phonemic equivalents must be emitted rapidly and in close succession. Even if phonemic equivalents are selected there is a limit on how rapidly they can follow each other. One does not arrive at speech by simply compressing the intervals between phonemes; the result of high compression will be an unintelligible blur, not the smooth flow of speech. In other words, the letters or phonemes must remain discrete and an upper limit on their rate of emission will again be set by the ear's temporal acuity. Each pronounced letter constitutes a syllable composed on the average of two elements or phonemes.

Each spelled word five letters long will thus have approximately ten elements. These, one may predict, will merge a high rates of spelling, say around 100 words a minute in theory, somewhat less in practice. This is certainly an advance on the direct translation, nonintegrating type of machine. There are perhaps advantages in this system from certain points of view which Dr. Metfessel describes.* We still cannot expect, however, optimal reading rates from the method.

Nor is this weakness solely due to the ear's limited temporal acuity; it is inherent in the display itself, as becomes clear if we consider how the visual system would handle such a display. As we know, the visual system can process printed text at a high rate; but could it do so if the text consisted of single letters briefly presented in rapid succession—the visual counterpart of spelled speech? Surely not. In short, if we want an auditory display comparable with the visual display of printed text we must ask more of our machine than that it should spell out loud.

To see how much more let us recall once again the sighted reader. If his first level of organization is the letter, his second is at least the syllable, if not the word or phrase. And if we assign this task to the machine—as it seems we must—we are choosing a recognition, syllable-or word-reading type of machine. That is to say, we are choosing a machine with a speech or at the very least a speech-like output.

The nature of the machine or machines that could yield such an output we will discuss below. Here we may remark that spoken language has one obvious advantage over spelled language: the phoneme elements that carry the information are encoded into higher order units—syllables. This is not to say that the minimal acoustic unit of speech is necessarily the syllable or that the phonemic segments interact and lose their identity in some higher Gestalt. On the contrary, the evidence is that speech—as produced and very possibly as perceived—may be described as the sum of independent, articulatory components at a level below even that of the phoneme (13). Nonetheless, the packaging of the phonemic segments into syllables in speech, as compared with their discrete delivery in spelled speech, has the consequence that whether we regard the phoneme or the syllable as the essential acoustic element, the number of elements per word is reduced. If we take the phoneme as our element, we have instead of approximately 10 elements per word as in spelled speech an average of around 3 or 4—a reduction by a factor close to 3. Thus we might

* See below, following paper.

expect that speech would become an unintelligible buzz at a rate of about 5 words a second or 300 words a minute. With the syllable as our element we would predict an even higher upper limit. In practice, around 200 words a minute is probably as fast a rate as one can comfortably listen to for any extended time. A reading machine with such an output will therefore still yield reading rates below those of sighted readers. But that is in the nature of the auditory stimulus and of the ear, and we may be confident that we shall find no auditory display with a higher rate of information transfer than speech.

Perhaps we may seem to have sidestepped the problem. We started by asking how a reading machine might best recover the acoustic flow that language lost when it went into print, and our answer turns out to be: by recovering the acoustic flow that language lost when it went into print. Obviously we would not be rash enough to make the suggestion were a reading machine with a speech or speech-like output not feasible. But recent work suggests that it is. Furthermore, a speech output has more advantages over nonspeech than a mere reduction in the number of discrete elements per word and the consequent increase in reading rate. Before we turn to the machine itself, perhaps it would be worth digressing for a moment to consider these advantages and some of the reasons for them.

First and foremost, of course, speech is speech—that is to say, a highly efficient auditory code with which the listener is already familiar. If the output of the machine is plain English—or plain any other language—what more can we ask? The listener is ready to use the machine with no more training than is needed to operate it.

But are the advantages of speech simply those of familiarity? We said above that we might be confident we would find no auditory display with a higher rate of information transfer than speech. The sounds of speech are indeed efficient vehicles of information, for they may not only pour from the speaker at a rate close to the limit of the human receiver's temporal resolving power, but they may also be accurately identified at that rate. Now it is well known that man's ability to discriminate (that is, to determine whether two stimuli are the same or different) is very good, but that his ability to identify in absolute terms (that is, to determine which stimulus it is) is relatively poor. But, as we have already remarked, much experimental work in recent years has shown that humans identify members of a set of complex or multidimensional stimuli with considerably more accuracy than they do members of a set of unidimentional stimuli. For example, Pollack and Ficks (17) found that they were able to transmit to their subjects as much as 6.9 bits of information per stimulus when their

stimuli were drawn from sets of auditory stimuli taking one of two values along each of eight dimensions. This is considerably more than the upper limit of 2.3 bits that Pollack (16) was able to transmit when stimuli were drawn from a set of auditory stimuli varying along the single dimension of pitch.

Some of our capacity for rapid and accurate identification of speech sounds may be due, then, to the complexity of these sounds. But we may note that Pollack and Fick's subjects did not work under pressure; in fact, they took as much time as they needed to identify the stimuli by marking a check list. Whether the rate of information transmission is increased as much as its quantity by complication of the stimulus we do not know; but we may reasonably suppose that the multidimensional nature of speech sounds is a necessary, if not a sufficient, condition of their being so precisely identifiable.

There are, in fact, good reasons for believing that there is more to the perception of speech than this. Several lines of evidence suggest that speech stimuli are perceived by reference to articulatory as well as acoustic dimensions. For example, the acoustic stimulus does not always display the invariance that our invariant perceptual response would lead us to expect. The acoustic cues for the perception of a given phoneme sometimes display abrupt discontinuities that are not reflected in the response: the response like the articulation remains unchanged. Again, there are other situations in which precisely the reverse occurs: discontinuities appear in both the perception and the articulation, but not in the acoustic stimulus. For example, many speech sounds are perceived categorically; that is to say, the change from the perception of one phoneme to the perception of another as we move smoothly along some acoustic continuum is not gradual, but abrupt. This categorical perception is paralleled by categorical articulation.

The explanation for these discrepancies may lie in a theory of speech perception that we have developed more fully elsewhere (12), namely, that the perception of speech is linked to the feedback from the speaker's own articulatory movement. According to this theory, the listener learns a connection between speech sounds and their appropriate articulations. In time, the articulatory movements (or, more likely, the corresponding neurological processes) come to mediate between the incoming acoustic stimulus and its ultimate perception. If this is so, we should expect that at points where articulation and sound divide, perception should follow articulation.

Thus there is a body of evidence—treated in detail elsewhere—sug-

gesting that speech sounds are perceived by reference to the articulatory movements that produce them and that this articulatory reference is important for their rapid, absolute identification.

Returning now to our theme, we may draw a conclusion. If our account of the perception of speech by reference to articulation is valid the advantages of accurate and rapid identification of sound elements will accrue to any output that may be articulated, but not to "unspeakable" stimuli that merely resemble speech acoustically. Thus even if the output of a reading machine is far from standard English—or any other known language—but is rather some strange yet pronounceable machine dialect, the listener will be able to follow it at a rapid rate.

Of course he will have to learn the dialect. But here again speech has an advantage over nonspeech: it is more easily learned. At the Haskins Laboratories some years ago experiments were conducted on the ease of learning real and simulated nonspeech outputs from various types of reading machines. "A synthetic pronounceable language (known as Wuhzi) based on a transliteration of written English which preserved the phonetic patterns of the words" was also used for purposes of comparison (2). Wuhzi was learned far more rapidly and yielded a markedly higher terminal performance than any of the nonspeech outputs. More recently, House et al. (8), have demonstrated a similar learning advantage for speech over nonspeech stimuli that resembled speech acoustically but were not pronounceable. In commenting on their results the authors say, ". . . an understanding of the process of speech perception cannot be achieved through experiments that study classical psychophysical responses to complex acoustic stimuli. Although speech stimuli are accepted by the peripheral auditory mechanism, their interpretation as linguistic events transfers their processing to some nonperipheral center where the detailed characteristics of the peripheral analysis are irrelevant."

To sum up, a speech or speech-like output from a reading machine has several advantages. First, compared with both nonspeech and spelled speech, speech reduces the number of discrete elements per word and permits the transmission of information at a rate that is both rapid and well within the resolving power of the ear. Second, compared with nonspeech, speech sounds are highly distinctive and may be identified with far greater accuracy and speed. Third, any coded output in a speech-like —that is, pronounceable—form is readily learned, and, of course, if the language is plain English the learning has already been done and the listener is at home from the start.

We have said that a reading machine with a spoken output may be feasible. Now we would like to support our statement by describing briefly two classes of machine that are being developed at the Haskins Laboratories. The first generates speech by rule from the individual letters of the text; the second compiles speech by sorting recordings of the individual words of the text into appropriate sequences. Both machines require character recognition units. The first will also require a logic unit containing rules for selecting the phonemic equivalents of letters appropriate to their context; the second will require a large random access memory or dictionary for storage of the prerecorded words. While these hardware requirements may well be met in the not too distant future, they are the province of the engineer. Here we are concerned solely with the spoken output of the machines.

First, let us consider the synthesis of speech by rules from a phonemic input. The basis for such a set of rules has been laid by a long series of researches in our Laboratories into the acoustic cues for the perception of speech. In this research spectrographic analyses of speech have been studied minutely and reduced to a skeletal form in which only their essential features are preserved. Figure 1 illustrates the procedure. The top line shows a print of the original spectrogram of the words, "Never kill a snake." The middle line shows a reduced version painted by hand. The bottom line shows the final simplified, painted version arrived at by trial and error in which the essential acoustic cues are preserved and even brought into relief. This pattern is effectively a set of graphic instructions for the frequency and amplitude display and its changes over time necessary to yield an intelligible version of the original sentence. If the pattern is reconverted into sound (by means of a photoelectric device known as the Pattern Playback [4]) such a version will be heard.

By extensive research over the last ten years, techniques have been developed for the hand painting of simplified spectrograms—or, in other words, for the synthesis of speech—without reference to original spectrograms, and these techniques have been explicitly formulated as a set of rules for the synthesis of speech by Liberman et al. (13). These rules are designed to be "few in number, simple in structure, and susceptible of mechanization" to convert a string of phonemes into reasonably intelligible speech at normal speaking rates. While this is not the place for an exhaustive discussion of the rules, it may be of interest for us briefly to examine their structure.

Earlier we remarked that the phonemic elements of speech, even though

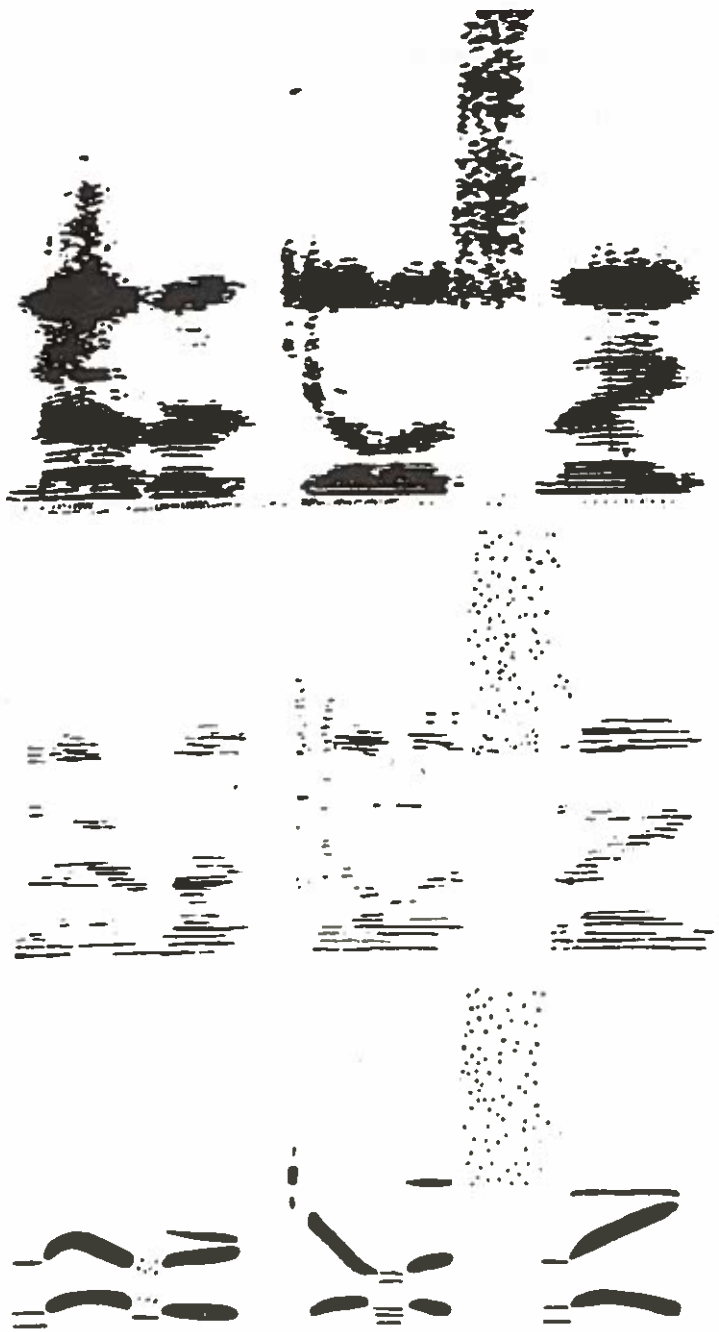

FIGURE 1 Spectrogram Patterns. Top: spectrogram of the words, "Never kill a snake." Middle: a reduced version of the spectrogram painted by hand. Bottom: a further simplification, with the principle features accentuated:

blended by the articulatory process into higher order syllabic units, do not lose their identity: they are independent or additive. This is not to say that the context in which a phoneme occurs is irrelevant. In applying the rules for the production of a given phoneme one must know the appropriate formant levels for adjacent phonemes so that the units may be sequentially combined into a smoothly flowing pattern. For some few phonemic combinations it is even necessary to write qualifications or "position modifiers" of the basic rules, but for the most part the rules may be so written that satisfactory sequential accommodations are achieved without modifiers. In principle the number of rules need scarcely be greater than the number of phonemes.

Furthermore, since the acoustic cues for the perception of phonemes fit readily into the linguist's articulatory classification, a further reduction in the ratio of rules to phonemes is achieved by writing the rules in terms of the subphonemic or articulatory dimensions of place, manner, and voicing. Statements must then, of course, be included within the subphonemic rules to permit their simultaneous combination. The rule for a stated class of phonemes thus consists of all the statements necessary to specify the acoustic cues on each of the subphonemic dimensions and to permit their simultaneous and sequential combination.

An example of synthesis by minimal rules is shown in Figure 2. The general comments about the relations between rules apply in the case of the phoneme /b/, for example, as follows. The set of rules for /b/ contains separate rules for the stop consonants /pbtdkg/, the labials /pbfvm/, and the voiced consonants /bdg/; certain statements in these rules must be combined simultaneously to specify, for example, the "silence"; and certain other statements must be combined sequentially with the rules for /æ/ and /z/, in particular those that generate formant transitions.

We have so far made no mention of prosodic features. In particular we have not mentioned stress, for which some provision in the rules is clearly essential. In natural speech stress is signaled by variations in one or more of the acoustic dimensions of fundamental frequency, intensity, and duration. In the minimal rules of Liberman et al. (13), only duration is used and only one stress modifier is included in the set of rules. An example of this is illustrated in Figure 2 in the position cell for /æ/.

To round out this brief discussion, Figure 3 gives an example of a longer utterance painted by rule. This was one of the earliest attempts and was painted for use on the Pattern Playback. Figure 4 shows a more elaborate example, also painted without reference to a spectrogram, this time for use with the Voback synthesizer (1). Note that this includes a pitch

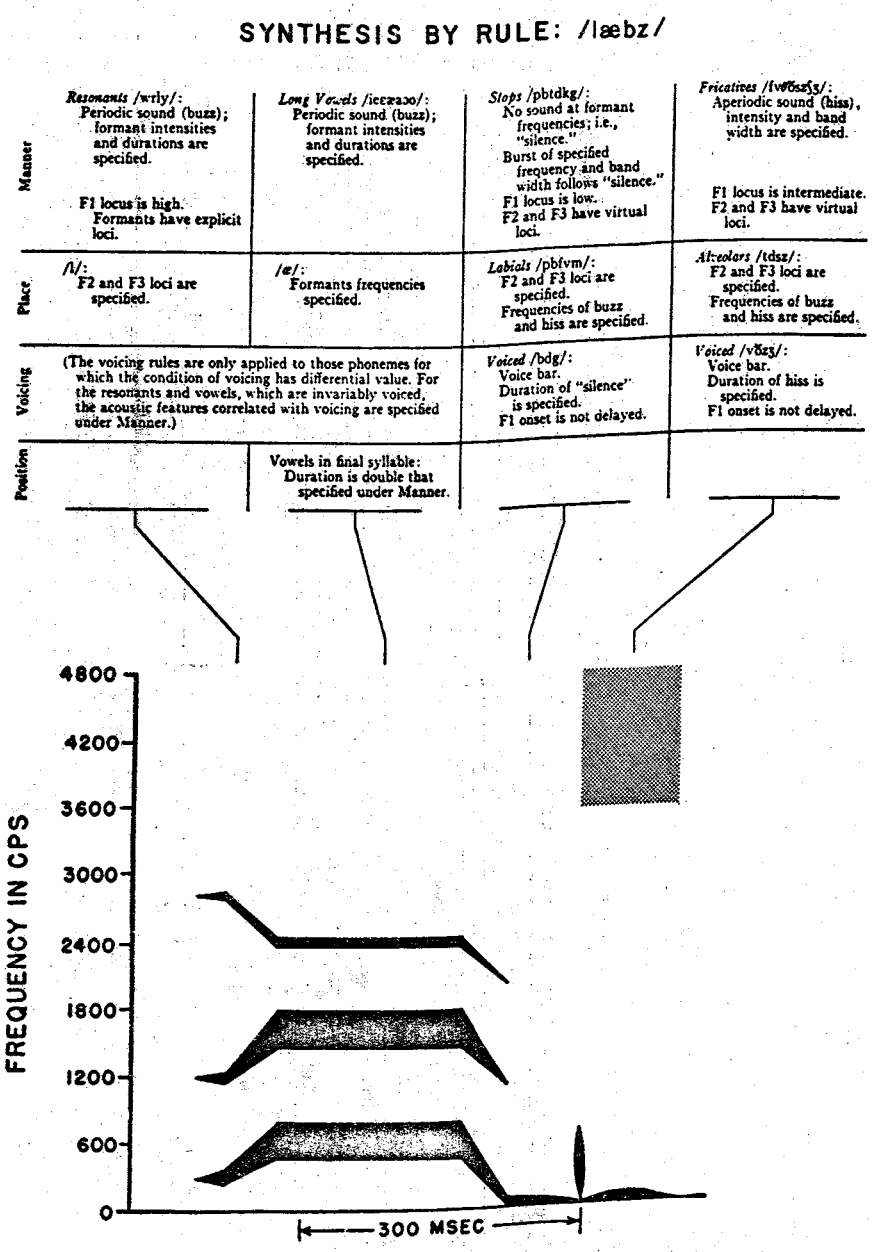

FIGURE 2 Word Synthesis. Top: rules for the synthesis of the word, "Labs" (/laebz/). Bottom: a reduced spectrogram, painted by hand according to the rules, for use on the Pattern Playback. For further explanation see text.

line, providing an added cue for intonation. A tape recording of these utterances has been made.

Obviously, such a set of rules achieves its parsimony and simplicity by reducing the acoustic specifications to those elements essential for recognition of the phoneme. Consequently the resulting speech is not entirely natural. It *is* readily intelligible and the rules do provide an explicit procedure for converting a phonemic transcription into control signals for a speech synthesizer. In practice, they have been used satisfactorily not only at Haskins Laboratories, but also at the University of Edinburgh (9). At the Bell Telephone Laboratories rules based on those described above have been used to program a digital computer, which then generated control signals and drove a simulated resonant speech synthesizer (10).

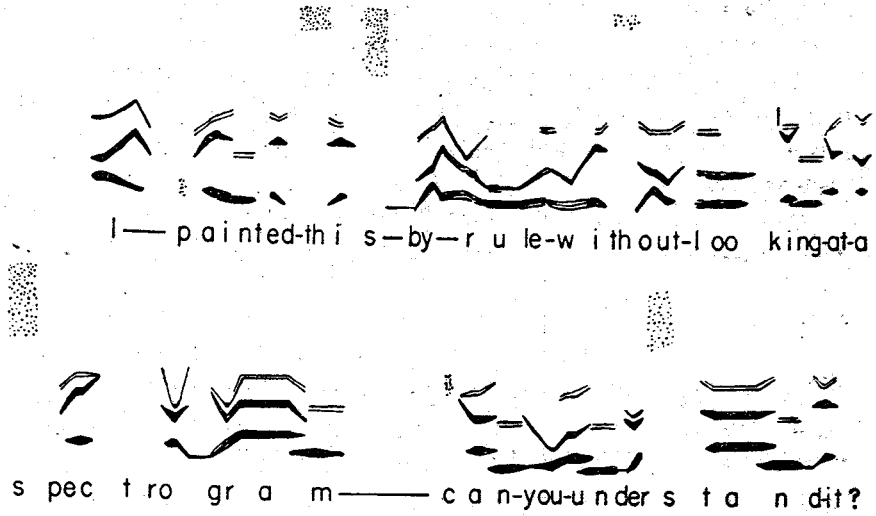

FIGURE 3 An Early Example of a Reduced Spectrogram Painted by Rule for Use with the Pattern Playback

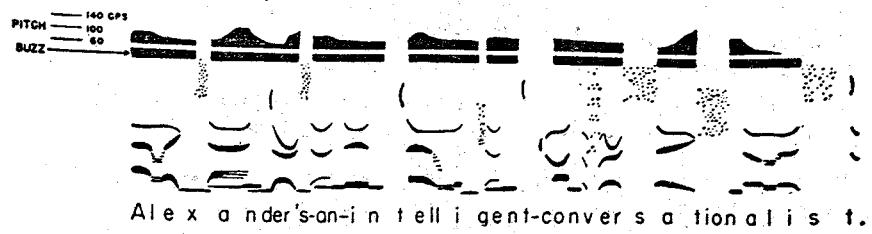

FIGURE 4 A More Elaborate Example of a Reduced Spectrogram, Painted by Rule for Use with the Voback Synthesizer (1). Note the pitch instructions which provide an added cue for intonation.

Finally, let us consider briefly the second method of generating speech under development at Haskins: speech by compilation. Here the first problem is the size of the speech segment to be used: phoneme, syllable, word, or phrase. The difficulties of using short segments have been discussed by several authors (7, 13, 15, 18). We have already remarked that the phonemic elements of speech, though linguistically discrete, are merged acoustically into higher order units. Just as we cannot specify clearly the temporal boundaries of the phoneme on the spectrogram, for example, so too we cannot synthesize the smooth flow of speech from discrete prerecorded phonemic elements. Peterson et al. (15) have attempted to by-pass this problem by using phonemic pairs or dyads that contain "parts of two phones with their mutual influence in the middle of the segment." However, the difficulties of matching the cut ends are not eliminated. Furthermore, Sivertsen (18) has shown that "the segment inventory becomes disproportionately large when the segments are not co-terminous with linguistic units." Similar problems are encountered with the syllable as a prerecorded element. It would seem in fact that words are the smallest elements we can hope to combine into reasonably natural cursive speech.

Yet the use of words is not without difficulties. Not least is the instrumental problem of how to store and retrieve the necessarily very large number of recordings, perhaps as many as 10 to 20 thousand in a satisfactory reading machine for the blind. By present-day techniques such a number would require a pair of very large disc or drum memory units. However, a smaller device with a vocabulary of some 7000 words is already being developed at Haskins Laboratories (3). This pilot Word Reading Machine will be used to evaluate the word compilation method for use in a reading machine for the blind.

The most important problem in developing the device has been how to record a single version of each word that will be acceptable for a variety of syntactical uses. The solution has been to assign each word to a grammatical class according to its most frequent usage, and to assign to each grammatical class an appropriate pattern of duration, intensity, and pitch change. A practised speaker is then instructed in terms that he can apply to the monitoring of his own output and, after some trial and error, a satisfactory recording is made (6). When the separate recordings are combined they yield an output that is highly intelligible at slow-to-average reading rates and that approximates the phrasal patterns of normal speech. Of course, the quality would be improved if the machine were provided with several versions of each word and with linguistic rules and logical circuitry to select the version most appropriate to the syntactical context. With the increase in quality would come, however, a perhaps disproportionate increase in complexity and cost.

In sum, we have argued that while there is obvious value in cheap and portable reading aids for the personal use of the blind, full access to our libraries can only be given by a high performance reading machine that will enable the blind to read at the comfortable rate of normal speech. The auditory output of the machine should be speech or at the very least speech-like. Such an output requires little or no training of the blind user. We may be confident for several reasons that we shall find no other auditory display capable of transmitting information either as rapidly or as accurately. Two methods of providing a speech display are currently being developed at the Haskins Laboratories: one method synthesizes speech from a phonemic input; the other compiles speech from a dictionary of prerecorded words. The program of development is still at too early a stage for a final choice between them to be made. The hardware required to link these outputs with the printed page is not yet available, but may well be so in the not too distant future. Granted this, we may reasonably hope that high performance reading machines with speech outputs will one day be installed in our public libraries and educational institutions.

REFERENCES

1. Borst, J. M., and F. S. Cooper, "Speech Research Devices Based on a Channel Vocoder," *J. Acoust. Soc. Amer.*, Vol. 29 (1957), p. 777.
2. Cooper, F. S., "Research on Reading Machines for the Blind," in P. A. Zahl (ed.) *Blindness.* Princeton: Princeton University Press, 1950.
3. Cooper, F. S., "Toward a High Performance Reading Machine for the Blind," in *Human Factors in Modern Technology.* New York: McGraw-Hill (in press)
4. Cooper, F. S., A. L. Liberman, and J. M. Borst, "The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech," *Proc. Nat. Acad. Sci.*, Vol. 37 (1951), pp. 318-328.
5. Eriksen, C. W. *Multidimensional Stimulus Differences and Accuracy of Discrimination.* Wright Air Development Center Tech. Rep. 54-165, June 1954.
6. Gaitenby, J., "Word-Reading Device: Experiments on the Transposability of Spoken Word," *J. Acoust. Soc. Amer.*, Vol. 33 (1961), p. 1664.
7. Harris, K., "Study of the Building Blocks of Speech," *J. Acoust. Soc. Amer.*, Vol. 25 (1953), pp. 962-969.
8. House, A. S., K. N. Stevens, T. T. Sandel, and J. B. Arnold, "On the Learning of Speechlike Vocabularies," *J. Verb. Learn. Verb. Behav.*, Vol. 1 (1962), pp. 133-143.
9. Ingemann, F., "Eight-Parameter Speech Synthesis," in *Progress Report.* Edinburgh: University of Edinburgh, 1960 (Phonetics Department).

10. Kelly, J. L., and L. J. Gerstman, "An Artificial Talker Driven from a Phonetic Input," *J. Acoust Soc. Amer.*, Vol. 33 (1961), p. 835.
11. Klemmer, E. T., and F. C. Frick, "Assimilation of Information from Dot and Matrix Patterns," *J. Exp. Psychol.*, Vol. 45 (1953), pp. 15-19.
12. Liberman, A. M., F. S. Cooper, K. S. Harris, and P. F. MacNeilage. "Motor Theory of Speech Perception." Preprint for Speech Communication Seminar, Stockholm, 1962.
13. Liberman, A. M., F. Ingemann, L. Lisker, P. Delattre, and F. S. Cooper, "Minimal Rules for Synthesizing Speech," *J. Acoust. Soc. Amer.*, Vol. 31 (1959), pp. 1490-1499.
14. Miller, G. A., "The Magical Number Seven, Plus-or-Minus Two, or, Some Limits on Our Capacity for Processing Information," *Psychol. Rev.*, Vol. 63 (1956), pp. 81-96.
15. Peterson, G., W. S-Y. Wang, and E. Sivertsen, "Segmentation Techniques in Speech Synthesis," *J. Acoust. Soc. Amer.*, Vol. 30 (1958), pp. 739-742.
16. Pollack, I., "The Information of Elementary Auditory Displays," *J. Acoust. Soc. Amer.*, Vol. 24 (1952), pp. 745-749.
17. Pollack, I., and L. Ficks, "Information of Elementary Multidimensional Auditory Displays," *J. Acoust. Soc. Amer.*, Vol. 26 (1954), pp. 155-158.
18. Sivertsen, E., "Segment Inventories for Speech Synthesis," *Lang. Speech*, Vol. 4 (1961), pp. 27-61.
19. Woodworth, R. S., and H. Schlosberg. *Experimental Psychology*. New York: Holt, Rinehart and Winston, 1954.