GUIDELINES FOR DEVELOPING AN ALPHABET TO SYNTHESIZE SPELLED SPEECH [a]

Milton Metfessel, Ph. D.

Constance Lovell, Ph. D.[b]

Metfessel Laboratories, Inc.
Los Angeles, California

FOREWORD

These guidelines for alphabet development form Part B of the summary report on "The Spelled-Speech Project of Research for Automatic Readers Usable by the Blind and the Blinded." In separate reports are presented: 1. a historical and conceptual review of the project and 2. the research done on training methods for use with spelled speech.

In 1969 Dr. Metfessel was a victim of California flood waters. Before his untimely death, a preliminary draft of this report, covering general features, had been written. That draft has been expanded by making use of his research notes. My participation in this phase of the research was sufficient to make feasible such supplementation. However, it was Dr. Metfessel who, over the years of the project, worked out the procedures reported in these guidelines.

Constance Lovell

I. INTRODUCTION

How can one develop a set of alphabet letter-sounds on tape for a reading machine for the blind or for other instruments requiring auditory readout? What needs to be done depends on the kind of spelled output to be synthesized: *discrete spelling,* in which there are gaps between letters, or *spelled speech,* in which the letter sounds are so close together that the combination tends to be perceived as a wordlike configuration.

Very simple procedures suffice to obtain an alphabet for discrete output. The most obvious way is to have someone speak the letters several times, running through the alphabet from A through Z. From a tape

---

[a] Part B of summary report on VA Contract V1005P–5321.

[b] Associate Professor, Department of Psychology, University of Southern California, Los Angeles, California 90007.

recording of such letter sounds, one then selects a clear representative of each letter of the alphabet, giving preference to short pronunciations and eliminating any that are markedly out of line in pitch, loudness, or tone quality. There is no necessity for more careful balancing of letters if they are to be presented discretely. Such an alphabet can be completed with only a few hours of work. Probably anyone familiar with the use of a tape recorder would have no difficulty in obtaining a satisfactory result.

An alphabet to synthesize discrete spelling is useful when individual components of a word need to be emphasized and when relatively slow rates (say 15 to 25 words per minute) are acceptable. However, for an automatic reader for the blind, an alphabet is needed that is not limited to slow speeds. Spelled speech up to 100 words per minute or more is the goal.

Obtaining an alphabet to synthesize this kind of output is vastly more complicated. In addition to being shorter, the letter sounds must be carefully balanced so they will go together smoothly in a manner approximating the phoneme flow of ordinary conversation. The problem of obtaining coalescent letters is increased by the need for enough pitch variation to avoid a monotone effect, which makes the spelled speech sound less natural. To achieve a satisfactory result it is necessary to take into account the voice used for the recording, the way in which the letter pronunciations are obtained, the characteristics of recording instruments, the relative frequency with which various letters occur in English words, and a myriad of other factors.

The guidelines for construction of spelled-speech alphabets presented in the following sections range from general working hypotheses to such specific details as location of letters on tape. The outgrowth of 12 years of work on alphabet construction, they are designed to enable a researcher to avoid blind-alley approaches and concentrate on procedures of promise. It is hoped that these suggestions will make it possible for anyone needing a spelled-speech alphabet to produce one in less than a year (as compared with the 5 years required for the first acceptable alphabet).

The guidelines are by no means a recipe for an "instant alphabet," nor do they provide a system of routine steps that will lead automatically to a successful result. Obtaining the first 20 letters or so is not difficult. However, it is misleading to judge the ease of the task by the speed with which this can be done. More time is likely to be spent on the last letter than on all the rest together. Each additional sound selected restricts markedly the field from which the remainder can be chosen. When the 26th letter is reached, it may be found that none of the pronunciations available will go well with all the earlier selections. It is then necessary to substitute alternative pronunciations for some of these, working back and forth until the desired fit is obtained.

II. GENERAL REQUIREMENTS

Tape Recorders and Their Use

In presenting the guidelines, it is assumed that those who use them have several tape recorders and that they are familiar with adjusting the machines to produce desired frequency-response patterns and to make duplicates that are equivalent. The instruments should provide for recording and duplicating at both 7½ and 15 in. per second. At least one recorder should permit accurate location of sounds on tape (within 1/16 in.).

Using a number of instruments in developing an alphabet is desirable because each machine places an imprint on material as it is recording. By successively duplicating material on different instruments, this effect can be utilized to increase the homogeneity of letter sounds and improve their coalescence. Special precautions must be taken to prevent the detrimental change in signal-noise ratio that usually occurs with successive duplication and makes for poor quality in the final product. This effect may be lessened by disconnecting the erase head during the duplication process. Another procedure sometimes helpful in taking out undesirable background noise is to duplicate a 15 i.p.s. recording by playing it at 7½ i.p.s. and using a speed of 7½ i.p.s. on the duplicating recorder.

Playback quality varies considerably with different instruments. Consequently, it is recommended that, throughout the process of developing an alphabet, it be tested on a model of the instrument for which it is needed.

An important aspect of the use of tape recorders is the location of sounds on tape. Suggestions concerning this are presented in Appendix A, together with material on splicing tape segments.

Listener Judgments Concerning Letter Sounds

In this work, thousands of decisions have to be made about sounds. For some purposes instruments are invaluable. (For example, checking similarity of V-U meter readings is used in deciding whether two duplicates are equivalent.) However, in determining whether a letter is appropriate for an alphabet, measurements of its physical characteristics are insufficient and at times misleading. The crucial question is: How is the sound perceived? Consequently, listener judgments play a key role. An illustration of this occurs in obtaining letters balanced with respect to loudness. An X and an M that give the same modulation reading on the meter are likely to be perceived by the listener as quite different in loudness. From the meter one reads the peak which each letter reaches, but apparently loudness is perceived in terms of the total temporal pattern, different for the two.

No matter how fine the auditory discrimination of the researcher working to develop an alphabet, it is important that he check his reactions

throughout the process by obtaining reports from others. Obtaining such judgments is complicated by the fact that the criterion of a "good" letter sound changes as the listener gains experience with spelled speech. Initially there may be confusion between some letters unless the differentiating phonemes are emphasized (for example, between N and M and between B and D). With practice, the listener learns to discriminate letters with small differences between them (the kind of sounds that facilitate coalescence). When this stage is reached, the "strong" letters originally needed are likely to seem unpleasantly stressed.

To take this factor into account, judgments concerning alphabet sounds should be made only by persons who have passed the initial stage of adjustment to spelled speech. Informal day-by-day checks can be done appropriately by those who are assisting with the research, provided there are controls to insure that they do not know the specific material in advance. This precaution is important, since expectation or preparatory "set" may have a marked effect on how material is heard.

Such checking needs to be supplemented by occasional testing with outside observers. On the project this was done with persons who were participating in studies of training methods. The sessions, extending over a number of weeks, gave individuals a chance to become accustomed to working with spelled speech. They performed in a setting in which their primary concern was getting the meaning of the material. This orientation encouraged their treating the spelled speech as they would ordinary spoken material, rather than applying to it standards more rigorous than would be used with natural speech production.

III. OBTAINING LETTER PRONUNCIATIONS

The first step of alphabet construction is to obtain a large number of letter pronunciations on tape. The goal is a supply of letters characterized by "unity within variety"—the former in the interest of coalescence and the latter for avoiding an unaccented montone effect.

General Nature of Recording Sessions

In obtaining pronunciations, emphasis is placed on staying as close as possible to the individual's customary manner of speaking. Requests for special kinds of pronunciation are kept minimal and, where introduced, are made as simple as possible. The hypothesis is that a person is more likely to produce a large number of letters that belong to the same universe of sound, if using long-standing speech habits, than if operating under instructions calling for conscious modification of speech. To facilitate natural speech production, the following are suggested:

1. Arrange the physical aspects for recording so that the person can speak comfortably into the microphone from a seated position. The aim

here is to facilitate maintenance of the same relationship between speaker and microphone as the session progresses.

2. Conduct a number of preliminary recording sessions to give the individuals a chance to become at ease in the situation.

3. Use recording materials in which the production of letter sounds is interspersed among sections of conventional speech. This tends to set natural bounds for the production of the letters.

4. Let the individual himself define what is a natural rate for him by the process of cycling rates. That is, ask him to alternate going through the material more rapidly than usual, at his normal rate, and more slowly than usual. With a given set of material, these rates may be cycled three or more times. The sections produced at fast and slow rates are not intended for alphabet selection, though available as a supplementary source of letters. They are used because they appear to help an individual determine what is a natural rate for him.

5. Do not give any instructions regarding loudness of speaking. If the person's natural voice level is above or below that desired, make adjustments in the placement of the microphone and/or the modulation setting of the recorder.

An appropriate modulation level is one in which the vowels are approximately in the center of the scale (say, between 45 and 75 percent modulation on recorders with V-U meters, like those on Berlant-Concertone and Ampex). With such a standard, some letters (e.g., S, F, and X) usually fall below this range and others (e.g., L and N) usually are above it. Natural variations occur from trial to trial for each letter and are desirable. During the course of a session, modifications in setting are made only if the modulation of the vowels shows marked and relatively lasting change.

6. Minimize instructions as to how the letter sounds are to be formed. It should be mentioned that in spelled speech W is shortened from "doubleyou" (dʌblɪu) to "dubyou" (dʌbɪu).[c]

Materials for Recording Sessions

Recording copy is planned so that letter pronunciations are spoken in a setting of conventional pronunciations. To obtain desired variations in pitch, the speaker is asked to produce the letters in small groups at different pitch levels, using descending order.

---

[c] In early work on the project the shorter contracted form "dubbie" (dʌ'bi) was used. It is readily perceived as W when followed by other letters. It was dropped in the interest of having a pronunciation that would sound more natural at the end of words and in isolation. Because W is used infrequently, rate is little affected by which of the two contractions is used.

One method of doing this is to use sentences such as the following:

```
                A
                 B
I think you are an  C
                   D
                    E

                                B
                                 C
Every girl on the beach deserved a  D
                                   E
                                    F

                C
                 D
Margie's car is a  E
                   F
                    G

            D
             E
Jay is OK if  F
               G
                H

                                E
                                 F
Most of my friends can teach   G
                                  H
                                   I
```

A small number of such sentences may be used, repeating them until all the alphabet letters have been produced. To provide change for the speaker during the session and to make possible one test of letters, recordings may also be made of longer series of sentences based on the same general plan. An illustrative set is given in Appendix B.

The following features of this method are to be noted:

1. The letters are cycled so that, in a complete set, each letter appears in each of the five positions. It is usual to find differences related to position. (For example, final letters are typically longer than the others.) Cycling of letter production results in numerous variations produced naturally by the speaker.

2. The letter sounds are spoken in alphabetical order in the interest of making the task as simple as possible. Supplementary orders can be introduced if study of the recordings of an individual shows difficulty in producing any two that occur successively. Such might well include a series in which a single letter is produced five times at the end of a sentence, as in the following:

```
                A
                 A
I think you are an  A
                   A
                    A
```

3. Regular use of the descending pitch order is also in the interest of keeping the task as simple as possible. The amount of pitch difference between the letters is left up to the speaker, and no attempt is made to control the melody of the preceding words.

Toward the end of the project, a second method of organizing recording materials was used that appeared promising. The speaker was asked to: 1. pronounce A four or five times in descending pitch order; 2. produce AA four or five times in descending pitch order; 3. proceed through the alphabet using the first two steps for each of the other letters. This procedure was interspersed between recordings of conventional pronunciation. It can very easily be used together with the materials previously described.

Selection of an Appropriate Voice

At the start it is desirable to make trial recordings of a number of voices. It is suggested that preference be given to individuals: 1. who speak clearly but in flowing rather than clipped fashion; 2. whose voices have uniqueness, such as some huskiness; 3. whose pitch level is low; and 4. whose natural rate of speaking is neither fast nor slow. Voice uniqueness seems to function as a common carrier factor that increases coalescence. With a voice of low pitch, some shortening of alphabet letters may be produced without a "Donald Duck effect" by increasing the playing speed of a tape that is being duplicated.

For each person, plan a recording session of about an hour, using the materials and procedures described in the preceding section. The most important information to be gained from this session concerns length of letter sounds. In earlier work on alphabet development, emphasis was given to having individuals speak the letters as rapidly as possible. However, experience since 1960 has revealed that it is much easier to obtain a set of coalescent letters if the individual naturally produces letters within certain maxima. The letters do not have to be the lengths required to achieve the final rate of spelled speech desired, for initially clear letters can be shortened by cutting sections from them.

To check this aspect, make a duplicate of the original tape, using a recording speed of 15 i.p.s. Use this duplicate to see if it is possible to locate readily identifiable letters that meet the following criteria:

E: not more than 2 in.
Other vowels: not more than 2½ in.
H, Y, and W: not more than 3 in.

If several individuals meet these criteria, consider the following in making a choice:

1. Ability to produce results in line with the directions. For example, one may take out the middle letter in each group of five in a series and

splice them together at short intervals. The closer to a monotone effect produced, the more consistent the person has been in maintaining "set."

2. Voice quality such that in pronouncing modifying consonants (e.g., "pl" in "pleases") the effect is a neutral sound, which is between a whisper and full voice but preferably closer to a whisper.

Alphabets have been developed with both male and female voices. It is hypothesized that the latter are likely to be more suitable for securing faster rates. In voices studied on the project, the characteristic formants of the sounds tended to have slightly more redundancy in those of females. This suggests that generally less would be lost by clipping segments from letters in a woman's voice.

Once an individual has been chosen for intensive work, additional sessions are conducted until a recording is obtained in which the cycling procedure has been followed with ease. With the materials described in the preceding section, it is probable that the letters of an alphabet will come from a single recording session. However, standardization from session to session of both tape and recorder settings is a desirable precaution. Furthermore, with some kinds of tape, variations both between and within reels are found. Spot checks with a frequency alignment tape (Ampex, Nortronics, etc.) are advisable.

IV. SELECTION AND DEVELOPMENT OF LETTERS

Preparation of Initial Duplicates

It is desirable to keep the original of the recording session intact for reference purposes. Consequently, the first step is to make a number of duplicates (5 or more), taking every precaution to insure that they are equivalent. In the material that follows, all letter lengths and other measures refer to tapes duplicated at 15 i.p.s.

Preparation of Working Duplicates

One of the initial duplicates is used for locating the beginning and end of each letter pronunciation in the natural-rate sections. Adjoining letters are then separated by splicing in about 7½ in. of leader or blank recording tape. If in playback it is discovered that part of a letter has been cut off, a copy of the same pronunciation from one of the equivalent duplicates can be substituted. Specific identifying information concerning each letter is written on the tape preceding it (e.g., B–1–2, to specify the B recorded on the first pronunciation of the second sentence).

In this process, letters of excessive length can be discarded (E more than 2 in., other vowels more than 2½ in., and other letters more than 3 in.). Pronunciations involving obvious errors, an extraneous noise during the letter sound, etc., can also be removed. If there is any doubt as to whether or not a letter is usable, it should be left in. The intent here

is simply to prevent having to spend subsequent playback time on material that is unsuitable. Especially to be avoided is an initial reduction of pronunciations by selecting those judged best when heard in isolation. Except for A, I, and O, the letters are going to be used in combination in spelled speech. Although those finally chosen should be readily identifiable when heard individually, they may be judged as less pleasing when so presented than pronunciations that do not coalesce as well.

Remaining material is then reordered, putting in the first section all segments of the tape containing the letters that are going to be considered, in the second section the parts containing conventional pronunciation, and in the final section the material pronounced at fast and slow rates. The time required for this reordering and for making a record of the contents of the changed tape is saved many times over in the reduction of the playback time that it makes possible. Several equivalent duplicates of this tape are then made, and it is these that form the working duplicates for initial selection of letters.

Selection in Terms of Pitch and Stress

The general plan is to choose most letters of average pitch and stress for the speaker and some that deviate in particular ways.

One may think of developing an alphabet with letters distributed in five pitch groups, approximately as follows:

Level	*Number of Letters*
1 (highest pitch)	3
2	5
3	10
4	5
5 (lowest pitch)	3

The goal is to have adjoining levels minimally discriminable—say, of such an order that those at a given level will be judged as higher than those at the next lower level by 70–80 percent of listeners.[d] The likelihood is that the original five pitch levels used in recording the materials will involve a greater range than desirable for the alphabet. The middle three groups will probably provide most of the letters finally selected.

No rigid prescription can be given with respect to which letters are to be at each level. There is latitude concerning the melody of an alphabet. However, experience with a number of alphabets suggests that:

1. E be at level 3.

[d] This may be checked by presenting the letters to a group of judges in a pair-comparison situation. In doing this it is important to control for order effect. Inflection within a letter sound may result in its being judged differently when presented before another letter than when presented after.

2. Other vowels be at level 3, but, if variation from this is necessitated by the pronunciations available, O and perhaps U be at level 2 and I be at level 4.
3. Levels 1 and 5 be utilized only for rare consonants (e.g., Z, V, B, J), with some at each level for purposes of variety.[e]
4. Level 4 rather than level 2 be used for consonants that occur more frequently at the ends of words than at the beginning (e.g., D and N), with the reverse for those that occur more frequently as initial letters (e.g., C and B).[e]

As with pitch, the relative stress of particular letters may be expected to vary from one alphabet to another. However, the following suggestions may be made:

1. That in selecting letters for consideration, work be concentrated on those of average stress or below.
2. That E be an understressed letter. This is strongly recommended.
3. That, if O and U are at pitch level 2, they be understressed and that, if I is at level 4, it be slightly understressed.
4. That, in general, letters at pitch levels 1 and 2 be below average in stress.

One starts the selection procedure by marking,[f] cutting, and splicing together (with a few inches of blank tape between), all letters spoken as the third in a group of five. Any that deviate clearly from a monotone effect or are above average in stress are removed. The letter sounds remaining form a source for the letters at pitch level 3. (This may have to be supplemented later by addition of pronunciations from the second or fourth positions in the groups.)

At this stage for the first time it is necessary to introduce a special marking procedure for the following: B, D, G, J, K, P, Q, T, and W. After the beginning of the sound has been found, the start of the letter is marked to leave 3/16–4/16 in. of blank tape ahead of the sound. If this is not done, identification of the letters is difficult, if not impossible, when they are combined with others.[g] The requisite amount of blank tape is considered an integral part of the letter length. Unless the cri-

[e] Tables showing the relative frequency of occurrence of letters in English and their frequency as initials and terminals are given in *Secret and Urgent* by Fletcher Pratt (Garden City, New York: Blue Ribbon Books, 1942).

[f] Whenever the term "marking" is used, it refers to making a complete identification of the letter as well as showing the bounds of the sound. If, as frequently occurs, there is not room on the tape itself for this, a separate written record is made.

[g] These letters start with short, low-energy crucial phonemes, and the spaces function to emphasize them. Duplication that builds up the intensity of the voiceless sounds to approximate the voiced sounds permits some reduction in the amount of space needed but does not eliminate it completely.

terion terminal phoneme is long, a similar space at the end may be needed for F, H, S, and X.

Double-Letter Test

For each pronunciation being considered, this test involves locating two copies on equilavent working duplicates and splicing them as close together as possible. Letters that do not form a smooth combination are eliminated from further consideration. It has been found that, if a letter does not go well with itself, it has small chance of coalescing with all the other letters. In essence, this check removes from consideration those pronunciations in which there is excessive rise or fall within the letter.

Eventually, this test must be made for all letters of the alphabet. However, initially it may be applied to, say, the 15 most frequently-used letters, those for which an eventual selection at pitch level 3 is most likely. As each additional letter is brought in for consideration, it is checked first with this test.

Selection of S and T Pronunciations

The phonetic equivalence of the carrier vowels of two groups of letters is used as a basis for locating pronunciations that will go together. The consonants, B, C, D, G, P, T, V, and Z all end with *ee* (i) ; and F, L, M, N, S, and X all begin with *eh* (ε) . At the start, work is done with T and S as representatives of these two groups. An S and a T are selected and spliced in the two combinations ST and TS. If they do not go together smoothly, other pronunciations are tried until a coalescent pair has been obtained. It is desirable to find at least five and preferably 10 such pairs of S and T pronunciations.

Selections of Vowels

After coalescent S-T pairs have been found, search is made for vowel sounds that will go with them. Selecting from five to 10 pronunciations of each of the vowels, they are tried out with the various S-T pairs, getting the combinations AS, SA, AT, TA, ES, SE, ET, TE, etc. Vowels that do not go well with one S-T pair may be satisfactory when used with another.

The five vowels that coalesce with an S-T pair are then checked to determine whether they go well together. This requires splicing all the two-vowel combinations, AE, EA, AI, IA, etc. In any instances where a combination sounds rough, search is made for replacement pronunciations that will go with the S and the T and with the other vowel sounds. The work involved in finding coalescent vowels is reduced by prior use of the S-T test. That is, if five vowel pronunciations all go smoothly with the same S and T pronunciations, when spliced before and after them, the probability is much increased that they will produce satisfactory combinations when spliced with each other.

The goal of this step is to obtain at least five sets of S, T, and the vowels within each of which all the letters are coalescent. To achieve these, it may be necessary to go back to the original source for additional pronunciations. For O, U, and I, one may also try out samples at different pitch levels.

Selection of Other Consonants

Each of the coalescent S-T pairs is used to provide a basis for selecting other consonants with the same carrier vowels. For each letter with final *ee,* pronunciations are spliced before and after an S until one providing smooth combinations is found. Since these new letters are all similar enough to the S so they fit with it, the chances are good that they are also similar enough to coalesce with each other and with the T and the vowels previously matched up with the S. This is checked by splicing a sample of bigrams, delaying the task of putting together a complete set until later in the process.

In similar fashion, choice is made of letters with initial *eh* carrier vowel on the basis of their going well with one of the T's that has been chosen.

The letters H, J, K, Q, R, W, and Y are tested with both S and T and with a vowel sound if one is indicated by the carrier vowel (e.g., Q with U).

Since some of these 19 letters are more likely to be problems than others, it is suggested that work be done first with R, L, W, Y, H, M, and N. Furthermore, with these letters, some additional testing is likely to be helpful. First they should be spliced before and after each of the vowels. It may be that the best choices that can be made do not seem as coalescent as the S-T-vowel set. Before making changes in any of the letters in these sets, however, it is advisable to try out some longer combinations. What may be judged as rough when heard in isolation and listened to with analytical concentration, may be readily acceptable when heard under more typical listening conditions.

In doing this extra testing, L and R may be treated together, trying out the following: 1. RAL, LAR, REL, LER, etc.: 2. RALREL, LARLER, etc.: and 3. some English words containing several letters. In similar fashion, H and Y and M and N may be treated as pairs.

Checks Pertaining to Discriminability

For F, H, S, and X it is helpful to make a specific phoneme test. With the first three, the carrier vowel is removed and the crucial phoneme is spliced in front of each of the vowels. If the result is a clear-cut "fay, fee, fie, foe, few" (feɪ, fi, faɪ, fou, fɪu), the pronunciation is probably a good one. If not, the likelihood is that the letter will cause problems of discrimination later. With X, each vowel is spliced in front of the final

phoneme, to see whether the result is "aches, eeks, ikes, okes, ukes" (eɪks, iks, aɪks, ouks, juks).

It was not found necessary to make this kind of test in selecting other letters. However, when it was tried later on some of the pronunciations with *ee* carrier vowels that had been selected, the results were satisfactory in terms of this kind of criterion. For example, the initial phoneme of the D, when spliced in front of the vowels, produced "day, dee, die, doe, due" (deɪ, di, daɪ, dou, dɪu).

At the conclusion of this process, one should have several pronunciations of each letter, including two or three complete alphabet sets that have passed the coalescence tests described. To find five such complete sets would be desirable, but unlikely.

Letter Length

Throughout the initial selection, priority is given to obtaining coalescent letters. Length is brought into the picture only by setting upper limits for the letters as spoken naturally. These requirements are such that a considerable number of good, clear pronunciations within the limits is likely.

Since the eventual goal is a set of considerably shorter letter sounds, the question may be asked: Why not start the process by eliminaing all letter sounds not short enough to produce the rate desired? This approach was used earlier in the project. The fact that the first acceptable alphabet was achieved with this emphasis indicates that it does not preclude obtaining a satisfactory set of letters.

However, obtaining coalescence is the major problem in alphabet construction. Consequently, it is recommended that one start by working for a coalescent alphabet with a rate of about 60 words per minute and with the plan of cutting letter lengths subsequently to produce higher desired rates.[h] Such an alphabet requires a mean letter length of 2 11/16 in.[i]

With the maximum letter-lengths given on page 160 for H, Y, W, and the vowels, a rate of 60 w.p.m. would require a mean letter length of

[h] Starting with a slower alphabet and shortening letters to produce a faster rate is suggested primarily as a means of helping with the problem of coalescence. However, it has the additional advantage of providing two versions of the same pronunciations. Training studies indicate the desirability, with some individuals, of working first with longer letters and moving to shorter versions as trainees progress.

[i] Variations in method of determining alphabet rate are discussed in *Spelled Speech: Historical and Conceptual Review,* Part A of the summary report on Contract V1005P–5321. The procedure used in obtaining this mean letter length was as follows: The figure of 4.5 for average length of English words (Fletcher Pratt, *op. cit.,* 252) was increased to 5.5 to allow for a space between words. With this base, a 60 w.p.m. alphabet would involve 330 units in 60 seconds, or mean time per unit of 0.18 second. With a tape recorded at 15 i.p.s., this would be a letter length of 2 11/16 in.

2 13/16 in. for the remaining letters. (Table 1 shows how this figure was determined, taking into account the relative frequency of occurrence of each letter.) It is likely that the original selection of letters will include

TABLE 1.—*Determination of Mean Letter Length Required in 60 w.p.m. Alphabet for Letters Other Than Those Previously Assigned Maximal Lengths*

Letter	Frequency in 1000 words [a]	Proportional occurrence	Expected frequency in 60 words [b]	Previously assigned length (inches)	Inches required in 1 minute
E	591	.107	35.3	2.0	70.6
T	473	.086	28.4		
A	368	.067	22.1	2.5	55.3
O	360	.065	21.4	2.5	53.5
N	320	.058	19.1		
R	308	.056	18.5		
I	286	.052	17.2	2.5	43.0
S	275	.050	16.5		
H	237	.043	14.2	3.0	42.6
D	171	.031	10.2		
L	153	.028	9.2		
F	132	.024	7.9		
C	124	.023	7.6		
M	114	.021	6.9		
U	111	.020	6.6	2.5	16.5
G	90	.016	5.3		
Y	89	.016	5.3	3.0	15.9
P	89	.016	5.3		
W	68	.012	4.0	3.0	12.0
B	65	.012	4.0		
V	41	.007	2.3		
K	19	.004	1.3		
X	7	.001	.3		
J	6	.001	.3		
Q	5	.001	.3		
Z	3	.001	.3		
Space	1000	.182	60.6	3.0	181.8
Total	5505 [b]	1.000	330.4 [b]		491.2

With recording speed of 15 i.p.s., 900 in. of tape are used per minute. The total of the final column above shows that 491.2 in. are needed for spaces, the letters H, Y, and W, and the vowels. (The total expected frequency of this group in 60 words is 186.7.) Thus, the amount of tape available for the occurrences of the remaining letters is 408.8 in. The 18 letters not previously assigned lengths have expected frequencies in 60 words that total to 143.7. Dividing 408.8 by 143.7 yields a value of 2.8 in., the mean letter length needed.

[a] The letter frequencies reported are from Fletcher Pratt, *op. cit.*, p. 252.

[b] Division of the total frequency for letters and spaces (5505) by 1000 gives an average word length of 5.5, including a space. Sixty words, thus, would be expected to average a total of 330 letters and spaces.

pronunciations short enough to produce this rate, and perhaps a somewhat higher one. However, reducing letter lengths by removing segments will undoubtedly be necessary if an alphabet at a considerably higher rate is desired. When this stage is reached, work focuses on one of the preliminary alphabets selected. The others are held in reserve in case problems develop subsequently that cannot be solved by change of one or two pronunciations.

In cutting letters, the general principle is to work toward final lengths in which the most frequently used sounds are the shortest. Some deviation from perfect negative correlation between length and frequency will probably be advisable in terms of other considerations. (For example, N and M may have to be longer than their frequency of occurrence would indicate to provide clear distinction between them.) However, the closer a perfect negative correlation can be approximated, the higher the maximum rate for the alphabet will be.

The specific segments that can be removed depend on how the individual speaker has formed the letter sounds. However, letters can be classified into four groups in terms of where removal of a segment is most likely to be possible:

1. At either end: E, H, X, and sometimes S and F.
2. At front of letter: A, I, J, K, L, O, R, U.
3. At end of letter: B, C, D, G, P, T, V, Z. Of these, T is the easiest with which to work. If shortening any of these letters does not produce the length desired, it may be helpful to try cuts in the center—with both the original pronunciation and with the version produced by removing a segment at the end.
4. In the middle of letter (between phonemes) : H, X, and W.

Table 2 gives a general idea of letter lengths to be sought. It presents the measurements for three different alphabets, with rates of 78, 90, and 115 words per minute, respectively. Because of variations from speaker to speaker in formation of letter sounds, it is not to be expected that these lengths would be duplicated in any other alphabets providing the same rates. It may be noted that among these three, there are variations in the relative lengths of the letters. Although most of the letter lengths in the 115 w.p.m. alphabet are shorter than those of the other two (as they would have to be to produce the faster rate), there are exceptions. The letters Q, W, and Z are just as long as those in the 78 w.p.m. alphabet; and the letters R, W, and Z are longer than in the 90 w.p.m. alphabet.

Once the maximum reduction of letter lengths has been made by removal of segments, some additional shortening may be obtained in duplication. Minor changes in speed may be made on some recorders by removing the felt pads opposite the heads. The same may be done by

TABLE 2.—*Letter Lengths in Three Alphabets (Measurement in inches)*

	Alphabet I (male voice, 78 w.p.m.)	Alphabet II (male voice, 90 w.p.m.)	Alphabet III (female voice, 115 w.p.m.)
A	2 1/8	1 11/16	1 1/2
B	2	1 7/8	1 3/4
C	2 3/4	2 3/8	1 7/8
D	2 1/2	1 11/16	1 7/16
E	1 3/4	1 7/16	7/8
F	2 1/2	2	2 1/8
G	2 1/4	2	1 9/16
H	2 1/2	2 1/4	1 13/16
I	1 7/8	1 13/16	1 3/8
J	2 7/8	2 1/2	2 1/4
K	2 1/2	2 1/2	1 3/4
L	2 1/8	2 1/4	2
M	2 1/2	2 1/16	1 1/2
N	2 1/4	1 9/16	1 11/16
O	2 1/4	2 1/4	1 1/4
P	2 1/4	2	1 9/16
Q	2 1/4	2 1/2	2 1/4
R	2 1/2	1 13/16	2 1/16
S	2 1/8	2 1/4	1 13/16
T	1 3/4	1 5/8	1 1/8
U	2 1/4	1 5/8	1 1/4
V	2 5/8	2 3/8	1 3/4
W	2 3/4	2 1/2	2 3/4
X	2 3/4	2 7/16	2 5/8
Y	2 7/8	2 1/2	1 3/4
Z	2 3/4	2 1/2	2 3/4

use of reels of different sizes or by manipulating the tape guides to increase or decrease tension. If the alphabet is played on an instrument adjusted to run at slightly faster than regular rate and the duplicate is made on one running at standard rate, the dupicate will contain slightly shorter versions of the letters at somewhat higher pitch.

Alphabets with Standard Letter Length

In developing alphabets for maximal spelled-speech rate, letter lengths vary from one to another. The design of some instruments for which spelled-speech output has been suggested calls for all letters to be the same length. An alphabet with varying letter lengths can be adapted for this use by leaving blank tape at the end of shorter letters. The soundless spaces thus introduced in some letter combinations are too short to make the spelling sound discrete, although the rhythm of the product is changed. With this kind of alteration, the rate obtainable with an alphabet is, of course, reduced.

An alternative would be to emphasize, in initial selection of letters that

tend to be long, the shortest pronunciations available. With this approach it is probable that much more time would be required to develop an alphabet satisfactory with respect to coalescence. Material synthesized with such an alphabet would be a mixture of letters spoken hurriedly and letters spoken naturally. Experience on the project with combinations of this kind led to the judgment that such output was less satisfactory. However, limitations imposed by the hardware of a reader might require either this approach or the more extreme one, currently under consideration,[j] of substituting other sounds for some, if not all, the letters. Both would involve a move away from natural-sounding spelling and would be expected to require longer learning time.

V. DEVELOPMENT OF A COMPATIBLE SPELLED-SPEECH ALPHABET

A major theme of these guidelines is that it is desirable to make use of the natural vocal habits of the speaker. The materials suggested for recording consist of a combination of ordinary speech and letter pronunciations. Emphasis is placed on setting up recording situations to minimize special instructions and to encourage a customary manner of speaking.

This approach is carried still further in developing a compatible spelled-speech alphabet. Such an alphabet is one in which the letters have met the criterion that they can be substituted in ordinary speech without detection.

Checking letters in this way requires a set of sentences in each of which there are at least two instances in which letter pronunciations form a word or a section of a word. The following sentence, e.g., contains five:

Most of my friends can teach chess.
O I N T S

The set of sentences in Appendix B was developed for this purpose. Except for W, each of the letters can be tested in a word or part of a word. For it, the most satisfactory situation devised to date involves use of "*dou*bly *beau*tiful." Since W is pronounced "dubyou" (dΛbɪu) in spelled speech, the letter can be divided and inserted as indicated.

The test is a limited one. It does not require that each letter fit naturally into every place in the set of sentences. In ordinary speech some words are emphasized and some are not produced fully. At the ends of sentences there is a tendency to prolong length and lower pitch. The test is limited to finding at least one niche for each letter.

---

j Glendon C. Smith and Hans A. Mauch, *Summary Report on the Development of a Reading Machine for the Blind* (Dayton, Ohio: Mauch Laboratories, Inc., 1971), p. 63.

In choosing words for the substituting of alphabet letters, the following points may be noted:

1. Words in the middle of sentences are more likely to be suitable than those at either beginning or end, which tend to vary more from one sentence to another.
2. Words that deviate markedly in either direction from average stress are unlikely to provide good places for substitution. Exceptions to this occur in the case of "if," "I'm," and "I'll." Imperfectly spoken, these words may provide excellent places for substituting F, L, and M, depending on the speaker.
3. If the section of natural speech and the spelled letter are of equivalent length, the best results are likely to be obtained by using a letter copy that is 10 percent less in modulation than the sound being replaced.

Results with this procedure may be tested objectively in a number of ways:

1. Individuals may be asked to write each sentence, providing check on relative clarity of synthesized and non-synthesized words.
2. They may be asked to indicate any words that sound unnatural to them.
3. They may be told that one of the words in a sentence is synthesized and asked to state which they think it is. The presence of at least two possibilities for substituting a letter is an important control with this procedure.

This approach to alphabet development was used with the last of the alphabets developed on the project. Much of the letter-by-letter testing outlined in the previous section was not needed because the process could be shortened by securing compatible letters at the beginning. However, this procedure is included in the guidelines as a supplement to the other procedure rather than as a supplanting method. Considerable experience with the "feel" of letters and their interrelations must be obtained from the start before heavy reliance can be placed on this approach. It may, however, be of value in checking initial selections prior to or concomitantly with testing in context.

VI. TESTING IN CONTEXT AND FINAL BALANCING

For evaluating the quality of spelled speech produced with an alphabet, a large sample of synthesized material is desirable—the more the better. If the instrument for which the alphabet is desired can be used to synthesize words (or if the process can be computerized), obtaining a large sample of material may not be a major task. On this project, such was not the case. Material was synthesized by making many equivalent

duplicates of an alphabet, and then marking and splicing letters together as needed.

When this method is used, it is helpful to make a complete set of bigrams. With all two-letter combinations available, the amount of splicing to produce words can be much reduced. In listening to the entire set, it is likely that some combinations will sound somewhat rough. However, before making any changes, it is desirable to synthesize longer combinations and to study how the letters sound in words put together in sentences and paragraphs. Evaluating letters on the basis of how they sound in longer combinations is in line with the fact that relatively few bigrams form English words.

The procedure used in building up a word library was to get as many words as possible with a minimum amount of splicing. Duplicates of the bigrams and alphabet letters were used to form three and four letter words. Duplicates of these in turn were made to obtain longer combinations. (In connection with this process it was discovered that the coalescence of an alphabet could be improved by successive duplication on different recorders.)

In the course of work with words and sentences, it may be obvious that a letter is a troublemaker. It may stand out in frequently used combinations, or it may echo in sentences, appearing to be present with or after other letters that follow it. In such instances, it may be necessary to go back to the original source for a more satisfactory pronunciation. However, before this is done, two other procedures are recommended. One is to make copies of the problem letter at different modulations and test these in the synthesized material. The second is to determine whether a change in the frequency pattern of the letter will make it more coalescent. In doing this it is suggested that the letter be reduplicated so as to emphasize low frequencies and reduce high frequencies.

In making judgments at this stage, it needs to be kept in mind that, because an alphabet is limited to a single pronunciation of each letter, the best conceivable cannot be expected to produce synthesized spelled speech either as smooth or as varied as rapid natural spelling. Some compromise with respect to both characteristics is necessary, since an alphabet providing maximum smoothness could be developed only by sacrificing variety.

What can be expected is that there be some melody and rhythm in the synthesized material and that there be no transitions that are jarring or letters that stand out as obtrusive. Experience on the project indicated that once an alphabet has been brought to this stage, it is appropriate to shift emphasis from letter selection to training. It probably is not possible to develop an alphabet in which all combinations will sound natural at first; nor is this necessary. Adaptation to unusual sounds occurs with experience—in listening to spelled speech as in listening to a foreign

language or to English spoken with a different accent. With repeated exposure, sounds originally regarded as strange and distracting eventually seem more natural; and awareness of the sounds as such is reduced.

The problem of coalescence has to be dealt with primarily through letter selection, to avoid sounds so jarring that adaptation would be difficult. Once this has been done, use can be made of the fact that, as the listener grows accustomed to the sounds, he contributes to the coalescence of the product.

APPENDIX A. LOCATING AND SPLICING LETTER SOUNDS

Marking Letters on Tape. For a person just starting to locate sounds on tape, the following are suggested:

1. Work with full-track recordings made at 15 i.p.s.

2. Start with a tape containing a number of pronunciations of the alphabet and sizable intervals between the letter sounds.

3. Make several equivalent duplicates of all the letters to permit more than one attempt at locating the same letter pronunciation. It is desirable to have all copies of the same pronunciation together, to facilitate comparison of various markings of the letter. This may be handled in the duplication process.

4. For locating sounds, work with a tape recorder placed upright, so that it is possible to grasp each tape reel with one hand and manipulate the reels manually when one is seated in front of the machine. The cover over the heads should be removed so that the tape is exposed and can be marked.

5. Manipulating the reels by hand, locate the sounds by rocking the tape back and forth under the playback head. High volume is helpful. Where the letters start and end is determined by joint use of three cues: auditory, visual (the reading on the V-U meter), and apparent pressure-kinesthetic (apparent change in feeling in the fingers at the dividing point between sound and no sound). Initiate the process with the letters B, D, P, and T, all of which have clear-cut starts. Special care is required in locating F, S, V, and Z because of their high-frequency components.

6. Mark the beginning and end of each letter. Various tape recorders differ in the ease and accuracy with which this can be done. On the project, most of the locating was done with Berlant-Concertone 30–30 instruments. With the start (or end) of a letter under the playback head, a mark was made a standard distance ahead of the point, and the tape was then advanced by that distance so that the point itself could be marked.

7. Splice in several inches of blank tape at the start and end of each letter. If part of the letter sound has been cut off at either end, it will be apparent when the result is played. To determine whether a letter is being marked longer than it actually is, mark and cut duplicate copies of the same pronunciation to shorter lengths and compare the results.

Splicing Letters. For this work it is recommended that use be made of a splicing bar (such as the Carson tape splicer), rather than one of the so-called "automatic" splicers. General instructions are provided with the instrument. However, some additional suggestions may be made:

1. It is desirable to demagnetize splicer and clamps, razor blades, and ballpoint pens used for marking at the start of each work period. This helps to prevent clicks at the splices.

2. Letter sounds are marked on the front (dull) side of the tape. If tapes are placed with this side up in the splicer for cutting purposes, it is then necessary to turn the tape over for affixing the splicing tape. This step can be eliminated by making a dot at the point to be cut with sufficient pressure so that its location can be determined accurately on the back of the tape.

3. The usual method of applying splicing tape is to place it across the joint from one side to the other of the magnetic tape, and to cut it in the splicer at both places using the guides provided. Half-inch splicing tape is generally used for this purpose. Experience on the project indicated that splices have less effect on playback results if 7/32 in. splicing tape is used and if it is placed in the splicing bar horizontally (in the same direction as the magnetic tape). A diagonal cut is made at one end of the splicing tape before it is put in the splicer. The splicing tape is then pressed down on the magnetic tape so that there is about ¼ in. (more rather than less) on each side of the joint. The second cut in the splicing tape is made by putting the edge of a razor blade at the desired place and pulling off the rest of the tape against it. The splicing tape is cut diagonally to facilitate its going under the playback head smoothly.

4. Once letters have been marked and cut, it is a great time-saver to be able to use them more than once, resplicing them into various different orders as appropriate. The fragility of tape limits the extent to which this can be done. It is of help in this situation if, when splices are taken apart, the original splicing tape is not removed completely. Instead, the splicing tape is left at the front of each letter and peeled from the end of the letter preceding it. (Use of 1½ mil. tape minimizes curling of the letter when the splicing material is pulled from it.) The splicing tape projecting from the front of a letter can then be slipped under the end of a new letter and pressed down firmly.

5. Especially to be avoided is a joint in which any of the white splicing tape is visible from the front of the recording tape. With repeated use of spliced material there is some tendency for splices to pull apart. This is increased if any of the adhesive surface of the splicing tape is exposed at the outset. At best it cannot be eliminated entirely. Consequently, duplicates rather than original spliced material should be used wherever possible.

APPENDIX B. TYPES OF SENTENCES USED FOR CHECKING ALPHABET LETTERS IN CONVENTIONAL SPEECH [a]

A There was only one daisy for each of the girls.
B The freshman team beat the varsity team.
C You can see the picture on the wall.
D The drive demands payment in dollars or work.
E Nothing pleases like high regard.
F The whole family left for the seashore.
G Margie's car is an F–85 or a Ford.
H He put each paycheck into his checking account.
I All the guys in the gang left for the beach.
J City Hall is adjacent to the State Building.
K She baked a cake for the party.
L The story I'm going to tell is not very long.
M Only one of the eminent people could pass the test.
N Ten stories up in Jay's office.
O There was only one car in the parking lot.
P The small group of people made a lot of noise.
Q I think you are a real expert on the subject.
R It is a smart idea to start early.
S Margie and Jake suggest that we buy a Ford.
T Most of my friends can teach chess.
U Very few of the students were present for the test.
V The mayor's veto ran into trouble.
W The view of the beach was doubly beautiful after the hazy weather was over.
X Everyone I know is an expert on baseball.
Y I understand why you bought that book.
Z If you take it easy, you'll do well.

[a] Each sentence contains a pronunciation of the letter listed at the left, plus at least one other letter pronunciation. If sentences are to be used to test whether alphabet letters can be substituted in them without detection, they are completed in each instance with a meaningful word. By substituting a group of five letter sounds, spoken in descending order, they can be used to give variety in the recording session.