Recent Improvements in Braille Transcription

Jonathan Allen and W. Terry Borroz, MIT

Abstract: Standard Grade II Braille uses 189 contractions to decrease the bulk of Braille text, and increase the speed of reading. The rules for the use of these contractions, however, are ambiguous and difficult to apply. A method is presented for detecting letter-string units which are functioning as linguistic units (morphs or clusters) in words. These units may then be contracted by either the Grade II system, or a more efficient procedure based on the frequency of occurrence of these units. Thus Braille can be encoded algorithmically by computer, and many of the "sequence" errors associated with other schemes are avoided.

KEY WORDS AND PHRASES: Braille Transcription, Linguistic units, Morphemic Analysis

CR CATEGORIES: 3.32, 3.34, 3.42, 3.63

INTRODUCTION

The problem of developing reading aids for the blind has typically been approached by converting the text message into a signal used to stimulate an alternate sensory modality, usually either auditory or tactile. Among these methods, tactile representations of letter signs have been used for at least six centuries. Many different codes have been used, and there has been a tendency, particularly during the present century, to make the code symbols and their relation to the original text increasingly abstract. The most popular tactile code for reading in use today was invented by Louis Braille in 1829, and the system still goes by his name. Although there are probably as many as 1,000,000 severely visually handicapped persons in the United States, only about 40,000 of them are Braille readers, while over 75,000 use talking books. The low use of Braille can be ascribed to many factors, which include the difficulty of learning Braille, the large bulk of Braille books, the time required to convert text to Braille (often several months), the cost of the process, the restricted number of texts available, and the extensive training (up to two years) required to produce qualified Braille transcribers. These considerations have led to the desire to automate the process of converting normal orthographic text to Braille, by using modern computational equipment and procedures. This overall task may be thought of as having three parts: input (converting the ink-print text into machine-readable form), code conversion (converting from normal letter codes to Braille codes), and output (embossing the Braille cells on stout paper). The input phase requires the ability to recognize characters of various font styles, and to represent in some formal manner equations, figures, and other special forms. Output requires the availability of a machine to emboss the dots of Braille codes on heavy paper. This has been done by modified teletypes and line printers, as well as specially designed embossers and Braille presses.

In this paper, we will be concerned with the code conversion process, although we do not imply that the input and output problems merit no further attention. We will describe the existing Braille code, develop a set of requirements for an ideal code, and propose a new code and transcription procedure which is designed to meet these requirements.

The output representation of the Braille code is based on the standard Braille cell, shown in Figure 1, which consists of two vertical columns, each having three possible embossed dots.

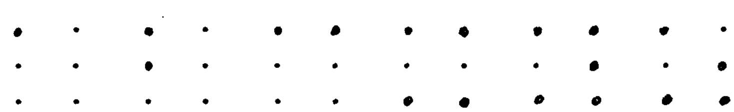
```
1 . . 4
2 . . 5
3 . . 6
```

Figure 1

Standard Braille Cell, with numbered dots

Thus there are 63 possible codes, not including the blank cell, which is used for space. There is no intentional relation between the arrangement of dots in a cell and the shape of the corresponding ink-print character.

Originally, Braille included only characters for each letter of the alphabet, numerals, and punctuation marks. This corresponds to Grade I Braille, where there is nearly a one-to-one correspondence between letters and Braille cells. In Figure 2, several single-cell letter codes are shown



A B C X Y Z

Figure 2

For Grade I Braille, there is clearly no difficulty in algorithmically transcribing text into Braille. While transcription is easy and the code is easy to learn, however, reading speed is slow, and the resulting bulk of Braille paper is large, so there has been understandable pressure to introduce new codes for frequent words and selected letter strings in order to increase the speed of reading and decrease the bulk of paper.

Since 1932, the most popular form of Braille has been Grade II, which utilizes 189 "contractions", some of which are shown in Figure 3. Grade II Braille uses all of the available 63 dot combinations, many of which have multiple meanings. In straight prose, this system uses only about 60% of the cells needed in Grade I Braille.

BRAILLE CODING DIFFICULTIES

Over the years a great deal of debate has centered on the units to be contracted, and the rules for specifying contractions. Various groups have argued for "sequence" rules, at one extreme, where a string of letters is always contracted, regardless of its environment, while others have insisted that contractions not be allowed across syllable boundaries. For instance, the "syllabification" forces would not allow the "ea" letter string in "react" to be contracted, whereas the "sequence" proponents would allow the contraction. It has not been possible to develop reading tests which would conclusively resolve this debate, with the result that current rules are often contradictory and lacking fundamental principles which would guide the selection of units to be contracted.

Nevertheless, it is possible to list a set of requirements for an ideal code:

1. The code should provide short contracted codes for frequently

occurring units.

2. The code must be readable. I.e., it must be easy to learn, the contractions must be easy to recognize, and reading speed should be high, with low fatigue.
3. It must be possible to transcribe text into Braille algorithmically, with as few exceptions as possible. Thus the rules must be explicit, unambiguous and consistent.

We now describe current difficulties with automatic transcription, and then show that word decomposition algorithms designed to reveal phonetically related units can remove many of these problems and provide a fundamentally sound approach to higher-level Braille translation.

The major difficulty with current "syllabification" Grade II Braille rules is that they are context-sensitive in an involved way.(1) Restrictions on the rules are based on position of a letter-sequence in a word, pronunciation of the letter string, and syllabification of the word. For example, "be", "con", and "dis" can be contracted only as syllables at the beginning of a word. Thus CONcept, but cone; DISturb, but disc, BErate, but bell. Note that in each case the contraction represents a prefix, even when it attaches onto a bound form, such as -cept and -turb. However, the rules also provide that the contraction COM may only be used at the beginning of a word, but it need not be a syllable. Thus COMe, but BEcome, and uncommittED. Final letter contractions (FULL, LESS) should be used in the middle or end of a word, but not at the beginning of a word. Nevertheless, a contraction must not be used where the usual Braille form of the base word would be altered by the addition of a prefix or suffix. Thus FUL is not contracted in "unfulfilled".

The desire to use many contractions has led to many instances of "correct" contractions which are obviously not related to the original meaning of the unit. Thus we have fiNESSed, fENCEd, and astrINGent. Further examples are shown in Figure 4. Notice that in every case, a contraction which is obviously legitimate linguistically in some contexts, is used in other places where the letter string which is contracted is <u>not</u> functioning as the linguistic unit. On the other hand, in many cases the letter strings to be contracted <u>are</u> legitimate linguistic units, as in careFUL and CONcept. It appears that the desire for linguistic correctness often bends to the goal of minimum number of Braille characters, but there is a limit, enforced by syllable boundaries, so that changEAble, miSHandle, miSTrust, and sweeTHeart are incorrect, even though each two-letter string does function as a unit in fEAst, SHape, STrike, and THirst, respectively.

It has proved to be extremely difficult to describe the "correct" Grade II contractions by rule, with the result that very elaborate conversion programs have been written which still fall short of producing official Grade II Braille.(2) There are even supposed semantic problems, as it is claimed that it must be known that "baroness" and "lioness" are feminine nouns, in which case NESS is not to be contracted. Finally, "do" is contracted to D when it is a verb, but not when it's a musical note, thus implying the necessity of syntactic analysis for proper conversion.

<u>LINGUISTIC UNITS IN WORDS</u>

The above examples show that Grade II Braille is based on the contraction of frequently occurring linguistic units of English, although these contractions are sometimes used beyond their original context. These units are of two types. Some are morphs, or basic lexical units of the language. These include prefixes (com-, be-, mini-); roots, both free and bound (snow, boat, house, -turb, -ceive);

derivational suffixes which affect the meaning of a word (-dom, -ness, -ship, -al) and inflectional suffixes, which affect the grammatical role of a word (-s, -ed, -ing). These morphs are the basic lexical units of English, and all words are made up from them. The two processes by which this occurs are compounding, where two roots are concatenated (house-boat, snowplow, uplift), and affixation, where either prefixes or suffixes are attached to roots (enable, receive, kind-nesses, affixation). Many of the units contracted in higher-level Braille codes correspond directly to these morphs, so that their proper detection is critical.

The second type of linguistic unit which is contracted is the cluster. There are both consonant clusters (st, sh, ch, chr, fth) and vowel clusters (ea, ou, ai). Many of these units are also contracted in Braille.

The existence of these linguistic units, and the native speaker's unerring recognition of them as functioning units, is obviously the main reason for selecting them as candidates for contraction. While the letter-string representations of these units occur in other contexts (e.g. mi<u>sh</u>andle vs. <u>sh</u>adow) it is the authors' belief that <u>if</u> these functioning morphs and clusters could be detected reliably from orthographic text, then there would be far less weight to the "sequence" arguments which would allow sh to be contracted in mi<u>sh</u>andle. That is, until the present, there has not been a principled technique for defining these units and detecting their presence, so that it is a matter of practical necessity to accept the "sequence" arguments in many cases. With this in mind, it is apparent that the basic requirements for a Braille code can be met if morphs and clusters can be detected when they are functioning as linguistic units, and that if the Grade II contractions are applied <u>only</u> to these detected units, then most of the common and difficult transcription errors will be avoided. To be sure, the resulting Braille will not be as fully contracted as the present rules indicate, but it is felt that this increased length (generally less than 10% of the official Grade II length) will be offset by the resulting ease of readability. The basic premise of this technique is to contract only true functioning linguistic units. Inasmuch as English speakers use these units continuously, the resulting Braille should be more natural and easier to learn and read. It is important to understand this point, since many "sequence" advocates claim that Braille readers readily adapt to "words that suffer through sequence", but it is readily conceded by most Braille readers that such non-linguistic use of contractions does slow down the reader when he meets the word for the first time or infrequently. It will be seen in the sequel that in some cases, proper recognition of the linguistic units leads to <u>more</u> contractions than the official rules would indicate. We propose, then, to replace the current ill-formed rules with an explicit procedure for detecting units to be contracted in <u>all</u> words. This scheme will nearly always find the true functioning linguistic units, at which point the existing set of contraction rules (or any revised set) can be applied. The trade between linguistic correctness and efficiency of representation is thus swung back toward proper recognition of units with an attendant increased length of encoded Braille. Thus HAD will be contracted, but not when it is embedded in sHADow. Similarly, fiNESSed would not be allowed, nor would THEsis. Whether or not this Braille is easier to learn or faster to read remains to be tested, but it is the authors' feeling that the new form will not be slower to read than official Grade II Braille.

Once the desirability of detecting functioning morphs is clear, there is the question as to how these units can be revealed algorithmically. Two ingredients are needed for this process: a dictionary of morphs and a procedure for discovering morphs in words. The latter set of rules is complicated by two factors. The first is that English vocalic suffixes (those which start with a vowel) often change the root word to which they attach. This can happen in three ways. A final silent "e" can be dropped, as choke + ing → choking; a consonant can be doubled, as pit + ed → pitted; or "y" can change to "i", as city + es → cities. Lee[3] has developed techniques for decomposing words into their constituent morphs while preserving the original non-mutated forms of the roots. Thus the three examples just cited are decomposed correctly, plus a vast number of additional words. More recently, however, Allen[4] has shown that the straightforward decomposition procedures proposed by Lee must be augmented by a set of selection rules. All possible decompositions are first obtained, and then the correct one is "selected" by these rules. For instance, scarcity → (scarce + ity / (scar + city, but the rules prefer affixation over compounding, so that scarce + ity is selected. Note that in this case, one decomposition involved the mutating vocalic suffix "ity", but the other case involves only the direct concatenation of two roots to form a compound. The selection rules are also useful when there are no mutating effects, such as resting → (rest + ing / (re + sting. In this case, inflectional affixation (-ing) is preferred over derivational affixation (re-), and the correct choice is selected. It turns out that a small number of these rules is capable of resolving a large set of ambiguities. Thus the recursive decomposition procedure introduced by Lee, coupled with the application of selectional rules to the resulting morph sequences, provides a powerful technique for revealing morphs.

These morph analysis procedures will not, of course, be useful unless they can be used in conjunction with a comprehensive morph lexicon which is capable of providing the content for a morph analysis of a large, representative corpus of English text. Such a dictionary has been obtained by Allen[4], who decomposed the entire Brown Corpus of 1,000,000 words of running text. The result is a morph lexicon of approximately 10,000 entries, sufficient to generate at least ten times that number of English words. It is important to realize that not only does this dictionary allow for the correct morphemic representation of a very large number of English words, but that the morphs in the lexicon are the basic atomic units of the language. Not only are they basic in the sense that all words are constructed from them, but they are also stable with time, so that very few new morphs are introduced in any appreciable span of time. Since new words are constantly being added to the language (e.g. "earthrise"), this stability factor means that the dictionary will not have to be continuously and substantially updated, as would a dictionary storing only complete words.

It is seen, then, that it is currently possible to decompose arbitrary English words into morphs, thus revealing their internal structure, and making it possible to avoid many sequence contractions (e.g. hothouse) as well as to reveal true units which should be contracted (knowledge/ful). Thus, for every morph which has a Braille contraction, the (Grade II) Braille code can be stored with the morph in the dictionary, so that once the structure is revealed, all whole-morph contractions can be applied at once. The detection of sub-morphemic letter-strings (bb, ea) must still be performed, but it is well to complete our discussion of the utility of morphemic analysis at this point.

In addition to supplying whole-morph contractions (as in knowledge/ful), the morphemic analysis also provides morph boundaries. Thus "hothouse" → hot + house, so that the medial "cluster" th is broken up. Neither "hot" nor "house" is contracted, but the derived boundary prevents the contraction rules from falsely contracting th when it is not a functioning cluster. These contraction rules must not be blind to the internal structure of the word, so that only linguistic and hence syllabic units are contracted.

Another feature of the morph lexicon is the frequency information associated with each morph, which is part of each lexical entry. The original Brown Corpus data provided the frequency, within the Corpus, of each of its words. Thus the word "the" occurred 69,971 times in the 1,000,000 word Corpus, and "blind" occurred 47 times. As each word was decomposed, two frequency counts were modified for each morph in the word. One count provides the number of distinct words in the Corpus which contain the given morph (call this N_w), while the second count gives the number of distinct words, times their respective Corpus frequencies, which contain the given morph (call this N_{cw}). Thus the decomposition of "blindfolded", with Corpus frequencies

blindfolded	1
blind	47
fold	7
-ed	(no frequency in original Corpus)

would add 1 to N_w and N_{cw} for each of the three constituent morphs. If the frequency of "blindfolded" had been say, 20, then 1 would have been added to N_w for each of the three morphs, but 20 would have been added to each N_{cw}.

From this description, it can be seen that two measures of morph frequency are available, based on actual occurrences in a large representative text sample. These figures provide the code designer with invaluable data on which to base the selection of morphs to be contracted. That is, not only can functioning morphs be accurately detected, but their frequencies are also known so that the savings due to the provision of a contraction for any of them can be computed straightforwardly. Suggestions for new contractions based on word frequencies have been made by Staack[5], but no data on morph occurrences has been available previously. In this paper, we have not suggested a revision of existing Grade II Braille contractions, but have dealt only with an initial system which detects linguistic units in words, but still uses only the existing set of contractions on these units. In this way any Braille reader trained to read Grade II code can easily use the linguistically correct form without the need to learn new contractions. In future studies, however, we do plan to suggest revised contractions, but only after some experience has been gained with the new linguistic Grade-II-compatible procedures.

It should be noted that if it is desired to contract letter strings in specified words or morphs which are not linguistic units, this can always be done by merely storing the Braille character string for that particular word. In other words, any number of exceptions to the underlying linguistic principles can be accommodated at the cost of increasing the size of the dictionary. The decomposition procedure does not have to be changed in any way when these exceptions are added to the lexicon, so that complete flexibility is retained. This ability to add special words may be particularly valuable for high-frequency proper names and technical terms for which contracted codes are desirable.

We have described how words may be analyzed into their constituent morphs, but clearly a complete morphemic representation will not always be possible. This is analogous to the situation of a native speaker when he neither recognizes a whole word nor observes any constituent morphs. The only recourse is to attempt to recognize functioning affixes, and then group the remaining letters into consonant and vowel clusters. Due to problems like swING and bleED, this is impossible in general when a complete morphemic decomposition is unavailable, so that such monomorphemic words must be entered directly in the lexicon. Nevertheless, there are rules that govern the sequence in which suffixes can attach to roots. Thus inflectional affixes such as "-ed" do not appear medially, as in firEDrake. Many other rules are known for a wide variety of suffixes, although contractions are not currently provided for many of these suffixes. Thus dictatorship → dict + ate + or + ship, even when "dict" is not available from the lexicon. This follows from the parts-of-speech relations between the affixes. For "-ship" makes nouns and attaches to nouns. Then "-or", which forms the "accepting" noun, requires a verb, and "-ate" makes verbs (as well as nouns and adjectives). In this way, it is frequently possible to recognize suffixes when the entire word cannot be represented as a string of dictionary entries. When this procedure succeeds, those suffixes which have Grade II contractions are immediately coded with the proper Braille cells, and morph boundaries are inserted at the suffix boundaries.

When a complete morphemic analysis is not obtained by decomposition, affixes, (if any) are first removed by the procedures described above. After these morphs are removed, it is assumed that the remaining letter string represents a root morph with no internal structure, i.e. it is monomorphemic. At this juncture, clusters can be detected. First, consonant clusters are detected, by scanning the string for maximal consonant sequences. This procedure insures that "chr" will be detected (if present) before "ch", so that clusters of all sizes will be revealed. After the consonants have been detected, vowel clusters form the remnant, and once again, the search is done for maximal units.

USE OF LINGUISTIC ANALYSIS FOR BRAILLE CODING

The result of all these linguistic analyses is two-fold. First, borders between all functioning units of a word are obtained. Thus:

houseboat - h|ou|s|e||b|oa|t

and contractions are then implied only internal to these boundaries. This may happen in two ways. Some morphs will be contracted in their entirety, such as "com-". In this case, it is not necessary to examine the cluster boundaries within the morph. In the remaining cases, morph boundaries (which are always syllable boundaries) block contractions across morphs, but the cluster boundaries provide the remaining sub-morphemic units which are available for contraction. For instance, dousing → douse + ing, so that we get one morph contraction (-ing) and one vowel cluster contraction.

The case of "dousing" requires further comment, since the decomposition has restored the final "silent e" of "douse". A Grade II analysis of this word would yield dousing, without the final e. Our procedure, running in the "compatible mode" can also neglect this "e" by simply performing the morphemic analysis to reveal "ing" and hence place a boundary before the "i" but then not restoring the "e" in the final output. Thus the "compatible" analysis would also yield dousing. This effect due to mutating suffixes, is much more widespread, and several examples are shown in Figure 5. Note that in some cases, the complete morphemic analysis results in more contractions, and hence fewer Braille cells in the resulting code. Thus we have:

rumbling	Grade II
rumbling	Grade II compatible
rumble ing	Morphemic

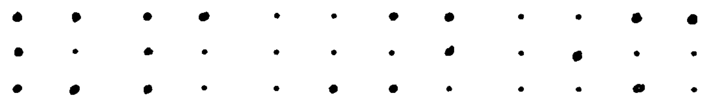
and

necessarily	Grade II
necessarily	Grade II compatible
necessary ly	Morphemic

At this point, it is clear that the combined use of morphemic analysis, suffix sequence tests, cluster detection, and Grade II contractions provides for a complete, unified procedure for Braille encoding which can be performed algorithmically. Further contractions can always be handled as lexical exceptions. Current problems such as fINe, gaSTiGHt, and dAR/EDevil are avoided. The price paid for these results includes the previously mentioned increased number of Braille characters (less than 10% increase over official Grade II Braille) and the necessity of providing for the morphemic decomposition procedure. The latter requires a dictionary for the morph lexicon, which can be expected to consume 32 to 64 K 18-bit words (and hence a disk store is necessary), plus the decomposition algorithm. The entire procedure can easily be managed on a mini-computer, and has been implemented on a DEC PDP-9. Other systems which do not provide morphemic analysis require very large exception lists, even though they cannot detect all of the desired syllable boundaries. Furthermore, the linguistic procedures take advantage of the basic structural properties of words, which do not change with time, and which preclude the necessity for the large exception lists required in other systems. New English words can almost always be represented in terms of existing morphs, whereas other systems may have to add the word to their exception lists. It is not possible to make a direct comparison between other systems which are intended to produce Grade II code and the present linguistically based system, since the end product is different. It is our conjecture, however, that the readability of the linguistic Braille will be high, and that the resulting Braille will be more accessible to a wider class of readers.

In order to provide readers with a clear and explicit view of the differences between these procedures, a sample text has been transcribed using three techniques for conversion. In Figure 6, the passage is converted according to the standard Grade II rules. In Figure 7, morph boundaries are detected, but mutations are not undone, so that the compatible (with Grade II) mode is produced. Finally, Figure 8 shows the same passage transcribed by applying Grade II contractions to a complete morphemic analysis, with mutations undone. The standard Grade II method requires 457 cells, the "compatible version 498 cells (+8.97% more than the standard method), and the morphemic process 495 cells (+8.32% more than the standard method). Large samples of text processed by each of these techniques are being produced, and will be evaluated by several Braille readers having varying amounts of training and experience. It should be noted, however, that there is no need for these readers to re-learn Braille, since only Grade II contractions are used. The only difference is that these contractions are now used only when they are justified linguistically, so that the new method is Grade II compatible.

It is felt that by contracting only those letter-strings which competent speakers of English readily recognize as units, the resulting Braille will be accessible to a larger community of readers, who can read it more naturally, with greater speed, and less training, than currently available automatically transcribed Braille.



AND PEOPLE -ATION MOTHER

Figure 3

Examples of Contracted Forms

ed: impede, breed, Eden, red
er: imperial, wither, permeable, peerless
and: incandescent, handle, offhand, abandon
in: finally, minus, lint, joined
the: theft, tithes, lather, Esther
one: hormone, gone, honest, money
of: offhand, sofa, aloof
en: men, penal, enamel, senile
ing: astringent, gingham, impinge
ance: chancellor, lancer

Figure 4

Standard Grade II Braille Contraction Examples

I. Final "e"			
	rumbling,	rumbling,	rumbleing
	Brailler,	Brailler,	Braille er
	conceived,	conceived,	conceive ed
II. "y" to "i"			
	necessarily,	necessarily,	necessary ly
III. Consonant doubling			
	running,	running,	runing
	ribb ing,	ribbing,	ribing
	sagged,	sagged,	saged
	saddest,	saddest,	sadest

Figure 5

Effect of Mutating Suffixes on Contractions
(Left-to-right order is: standard Braille, "compatible" Braille, Morphemic Braille)

|Four members of the ||MIT |Symphony |Orchestra are shown with four rare |Wagnerian tubas they learned to play in preparation for performances of |Bruckner's |7th Symphony on the Orchestra's recent spring tour. |Bruckner's symphony is one of very few compositions that use the tubas, which were invented by Wagner for his |Ring operas. |The instruments, two tenor and two bass, have a narrower bore than the conventional contrabass tuba (one of which is shown at center for comparison) and are equipped with a funnelshaped mouthpiece similar to those of |French horns. |This is one of less than a dozen sets ...

Figure 6

100 word text, showing standard Grade II Braille contractions

[457 cells: Underline => contraction, where number beneath equals the number of cells if not equal to one; vertical line => composition cell (capital sign, number sign, etc); ⌒ => no space between two words in Braille text.]

|Four members of the ||MIT |Symphony |Orchestra are shown with four rare |Wagnerian tubas they learned to play in preparation for performances of |Bruckner's |7th |Symphony on the |Orchestra's recent spring tour |Bruckner's symphony is one of very few compositions that use the tubas, which were invented by |Wagner for his |Ring operas. |The instruments, two tenor and two bass, have a narrower bore than the conventional contrabass tuba (one of which is shown at center for comparison) and are equipped with a funnel-shaped mouthpiece similar to those of |French horns. |This is one of less than a dozen sets...

Figure 7

100 word text, showing Grade II "compatible" contractions

[498 cells: Underline => contraction, where number beneath equals the number of cells if not equal to one; vertical line => composition cell (capital sign, number sign, etc.); ⌒ => no space between two words in Braille text.]

|Four members of the ||MIT |Symphony |Orchestra are shown with four rare |Wagnerian tubas they learned to play in preparation for performancees of |Bruckner's |7th |Symphony on the |Orchestra's recent spring tour. |Bruckner's symphony is one of very few compositions that use the tubas, which were invented by |Wagner for his Ring operas. The instruments, two tenor and two bass, have a narrower bore than the conventional contrabass tuba (one of which is shown at center for comparison) and are equiped with a funnel-shapeed mouth piece similar to those of French horns. |This is one of less than a dozen sets ...

Figure 8

100 word text, showing Morphemic Analysis of contractions

[495 cells: Underline => contraction, where number beneath equals the number of cells if not equal to one; vertical line => composition cell (capital sign, number sign, etc.) ⌒ => no space between two words in Braille text.]

This work was supported in part by the Joint Services Electronics Programs (Contract DAAB07-71-C-0300) and the National Institutes of Health Grant 5P01GM14940-06

REFERENCES

1. American Printing House for the Blind, English Braille, American Edition, 1959 (Revised 1968) Louisville, Kentucky

2. Gerhart, W.R., Millen, J.K., and Sullivan, J.E. Dotsys III: A Portable Program for Grade II Braille Translation Mitre Corp., May 1971

3. Lee, F.F. "Machine-to-Man Communication by Speech Part 1: Generation of Segmental Phonemes from Text." Proc. 1968 Spring Joint Computer Conference

4. Allen, J. "Development of a Dictionary for Speech Synthesis." Abstract, 80th meeting of the Acoustical Society of America, Nov. 4, 1970

5. Staack, G.F. "A Study of Braille Code Revisions" unpublished M.S. Thesis, MIT Dept. of Mech. Eng., August 1962