Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics

HONEY S. ELOVITZ, RODNEY JOHNSON, ASTRID McHUGH, AND JOHN E. SHORE, MEMBER, IEEE

Abstract—Speech synthesizers for computer voice output are most useful when not restricted to a prestored vocabulary. The simplest approach to unrestricted text-to-speech translation uses a small set of letter-to-sound rules, each specifying a pronunciation for one or more letters in some context. Unless this approach yields sufficient intelligibility, routine addition of text-to-speech translation to computer systems is unlikely, since more elaborate approaches, embodying large pronunciation dictionaries or linguistic analysis, require too much of the available computing resources.

The work here described demonstrates the practicality of routine text-to-speech translation. A set of 329 letter-to-sound rules has been developed. These translate English text into the international phonetic alphabet (IPA), producing correct pronunciations for approximately 90 percent of the words, or nearly 97 percent of the phonemes, in an average text sample. Most of the remaining words have single errors easily correctable by the listener. Another set of rules translates IPA into the phonetic coding for a particular commercial speech synthesizer.

This report describes the technical approach used and the support hardware and software developed. It gives overall performance figures, detailed statistics showing the importance of each rule, and listings of a translation program and another used in rule development.

I. Introduction

PHONETICALLY programmable speech synthesizers of reasonable intelligibility are commercially available today for a few thousand dollars. Such devices have stimulated widespread interest in computer voice output for various civilian and military applications. Among the most promising of these are the following:

1) transmitting information from English-language data bases to remote locations by telephone;

2) spoken output from reading machines for the blind;

3) communication with busy operators of computer-controlled systems who have to give most of their visual attention to complicated displays.

A further impetus to DoD interest is resulting from the development of narrow-band, digital voice transmission systems, such as the Naval Research Laboratory (NRL) linear predictive coder [1], and the likelihood of their widespread future use. These speech transmission systems include a synthesizer that could also be used for computer voice output.

The potential utility of computer-controlled speech synthesizers is greatly enhanced if the speech is not restricted to a prestored vocabulary. Among the numerous approaches to providing such unrestricted text-to-speech translation, the simplest is to use a small set of letter-to-sound rules to guess at the pronunciation of any word. Each rule specifies a phonetic correspondence to one or more letters. The letter's context is often used to determine which rule should be applied. An example is the well-known elementary school rule: "when two vowels go walking, the first one does the talking," which indicates that when one vowel is followed by another, the first is transcribed into the long vowel phoneme, whereas the second vowel is silent and receives no phonetic symbol.

A more elaborate approach, and one requiring much more storage, uses a large pronunciation dictionary supplement by various sets of rules. Words are isolated from the text and looked up in the dictionary. If the lookup fails, various rules are used to break the word into constituent parts for which there are dictionary entries. If all else fails, letter-to-sound rules are used to guess at the pronunciation. A further step in elaboration adds syntactic analysis of sentences to determine the part of speech of each word. This resolves the pronunciation ambiguities of words like *approximate* (adjective or verb?) and *house* (noun or verb?), although it falls short of deciding whether *unionized* refers to unions or ionization.

Text-to-speech systems have been built ranging in complexity from letter-to-sound rule systems to dictionary-lookup systems with syntactic analysis. Among the former, we mention those developed by Ainsworth [2] at the University of Keele and by McIlroy [3] at Bell Laboratories. The latter are exemplified by the system developed at Massachusetts Institute of Technology by Allen and Lee [4]-[8].

In order to be attractive as a routine addition to computer systems, text-to-speech translation cannot require a large fraction of the available computational resources. This constraint, which is particularly strong for real-time, military systems, precludes approaches that embody large pronouncing dictionaries or linguistic analysis programs. Thus, routine use of text-to-speech translation is likely only if sufficient intelligibility can be attained with a limited set of letter-to-sound rules.

In this paper, we report on work that has demonstrated the practicality of routine text-to-speech translation. We have developed a set of 329 letter-to-sound rules that translate English text into the international phonetic alphabet (IPA). Using the 50 000-word *Standard Corpus of Present-Day Edited American English* (Providence, RI: Brown Univ. Press) ("Brown Corpus") [9], we have determined that the rules will produce correct[1] pronunciations for approximately 90 percent of the words, or nearly 97 percent of the phonemes, in an average

Manuscript received November 7, 1975; revised August 10, 1976.

The authors are with the Naval Research Laboratory, Washington, DC 20375.

[1] The criterion for correctness is explicitly described in Section II-D.

sample of English text. The remaining words typically have single errors that in most cases can be corrected easily by the listener. A separate set of rules was developed to translate from IPA into a phonetic encoding compatible with a particular commercial speech synthesizer (Federal Screw Works Votrax VS-6).

The technical approach used in the NRL system is described in Section II as is the support software that we developed. Our results are summarized in Section III. Together with overall performance figures, we give detailed statistics that show the importance of each rule. Our conclusions and our plans for future work are discussed in Section IV.

II. The NRL System

A. The System

As discussed in Section I, the NRL system is designed to test the conjecture that acceptable intelligibility can be obtained with a limited set of letter-to-sound rules. Previous letter-to-sound rule systems, including those already mentioned [2], [3], had indeed already prepared us to believe this. None of them, however, for one reason or another, entirely satisfied our needs. One reason was lack of portability: either the program was written in assembly code for a particular computer, or it was designed around a particular hardware speech synthesizer. Sometimes performance measures were inexplicit enough that it was difficult to compare a system with other systems. A sentence like " the program translated *x* percent of the words acceptably" is uninformative unless one know: 1) what sample of words was being translated and 2) where "acceptably" falls in the range between "perfectly" and "comprehensibly to experienced listeners."

We imposed four requirements:

1) The implementation must be straightforward, requiring little space for the program and none for large dictionaries.

2) The system should not be tied to a particular hardware synthesizer.

3) The translation rules must be easily modifiable, both to allow for development and improvement of the rules and to permit the system to be tailored to a variety of special applications.

4) There should be an objective measure of the system's performance.

The NRL system is simpler than either Ainsworth's or McIlroy's in that it involves fewer specially coded preprocessing steps before the application of the rules. Thus, Ainsworth's initial pass for segmenting the input into breath groups [2] is absent; rather, we include rules that convert punctuation into pauses of various lengths. McIlroy's final-s stripping and ie-to-y conversion [3] are also absent, as is his look-up in an exceptions dictionary. Instead of a separate exceptions dictionary, we have included, for each word needing individual treatment, a rule giving its correct pronunciation. These rules are treated in no way differently from the rest; they make up about a sixth of the full set. The NRL system, like Ainsworth's, but unlike McIlroy's, does no fewriting of the input string and produces IPA as the output of the rules. The decision to use IPA was due to our desire not to be tied to a particular synthesizer; the text-to-phonetics information is contained in device-independent rules, and only the more direct phonetics-to-synthesizer rules need to be changed when it is desired to change to a new synthesizer. Likewise, for the sake of device-independence, the translation error scores we will present are based on the IPA output and so reflect only the performance of the pass-one rules. This is not to imply that a correct phonemic transcription is sufficient for high-quality output. The pass-two rules and the synthesizer hardware must together produce the correct allophone for each phoneme in context. Stress, rhythm, and inflection are likewise important.

Our work so far has used a commercial speech synthesizer, a Federal Screw Works Votrax VS-6 audio response unit. It can produce 63 different basic speech sounds ("phonemes") at 4 different pitch levels (inflections) and string them together to form continuous speech. The Votrax "phonemes" do not correspond exactly to the phonemes of English, but one can set up a fairly straightforward mapping from a phonemic transcription to Votrax codes. Some context-dependent rules are included, and the synthesizer itself takes context into account in pronouncing its "phonemes."

We used the synthesizer with a system of support devices that provides for convenient input, output, and manipulation of phonetic texts. These are described in [10].

B. Support Software

Because we require a convenient means of changing the rules in the course of their development, we have not immediately proceeded to a hand-coded system, like Ainsworth's, which incorporates the rules in the form of assembly code [11]. Among the research tools we have developed is a translation program in Snobol which contains the rules as a text string easily modifiable even by someone with no knowledge of Snobol. This translation program, TRANS, accepts text, applies the translation rules, and returns the translated results. The complete translation from English to Votrax codes may be requested, or the English-to-IPA or IPA-to-Votrax pass may be requested separately. This program is not, of course, the small, economical implementation evisioned for applications; we have paid in size and run time for convenience in experimentation.

The rules are kept in character strings in a form easy for human beings to read and write. Each rule has the form

$$A[B]C=D,$$

essentially the same form as Ainsworth's. The meaning is "The character string B, occurring with left context A and right context C, gets the pronunciation D."

D consists of IPA symbols—or rather a latin-letter representation of IPA to cater to computer character sets (see Table I). B is a letter or text fragment to be translated. A and C are patterns; like B, they may be strings of letters and other characters, but some special symbols denote classes of strings like "voiced consonant," "vowel cluster," etc. Table II lists the symbols that have such special interpretations. Blanks are significant, as they identify the beginnings and ends of words.

For example, a typical rule is

TABLE I
LATIN-LETTER REPRESENTATION OF IPA

Standard IPA	Representation	Example	Standard IPA	Representation	Example
i	IY	beet	g	G	goat
ɪ	IH	bit	f	F	fault
e	EY	gate	v	V	vault
ɛ	EH	get	θ	TH	ether
æ	AE	fat	ð	DH	either
a	AA	father	s	S	sue
ɔ	AO	lawn	z	Z	zoo
o	OW	lone	ʃ	SH	leash
U	UH	full	ʒ	ZH	leisure
u	UW	fool	h	HH	how
ɝ, ɚ	ER	murder	m	M	sum
ə	AX	about	n	N	sun
ʌ	AH	but	ŋ	NX	sung
aɪ	AY	hide	l	L	laugh
aU	AW	how	w	W	wear
ɔɪ	OY	toy	j	Y	young
p	P	pack	r	R	rate
b	B	back	tʃ	CH	char
t	T	time	dʒ	JH	jar
d	D	dime	hw	WH	where
k	K	coat			

TABLE II
SPECIAL SYMBOLS APPEARING IN THE ENGLISH TO IPA TRANSLATION RULES

Symbol	Meaning
#	One or more vowels*
•	One of B, D, V, G, J, L, M, N, R, W, Z: a voiced consonant
%	One of ER, E, ES, ED, ING, ELY: a suffix
&	One of S, C, G, Z, X, J, CH, SH: a sibilant
@	One of T, S, R, D, L, Z, N, J, TH, CH, SH: a consonant influencing the sound of following u (cf. rule, mule)
^	One consonant**
+	One of E, I, Y: a front vowel
:	Zero or more consonants

* Vowels are A,E,I,O,U,Y.
**Consonants are B,C,D,F,G,H,J,K,L,M,N,P,Q,R,S,T,V,W,X,Z.

'#C[O]M=/AA/',

which means that an O after an initial C and before an M gets the pronunication /a/ –the *a*-sound in *father*. Another example is

'#:[E]#=/IY/'.

Here the colon denotes any sequence of 0 or more consonants, so the rule says final E, if the only vowel in a word, gets the long-*e* sound /i/ of *be* and *she*.

The translation algorithm scans input text from left to right and, for each character scanned, sequentially searches the rules pertinent to that character until it finds one whose left-hand side matches the text at the correct position. It outputs the right-hand side, passes over the characters bracketed in the rule, and resumes the scan with the next character of text. The input string is never altered.

As an illustration of the operation of the algorithm, here is a worked example: the translation of RATIO using the English-to-IPA rules, which may be found in Table VIII.

To the left of the first character, R, the program adds a blank to delimit the word, and the scan starts with the R, as we indicate with a pointer: ↑RATIO. The program searches the R-rules—the rules with R as the first character between brackets. The first R-rule, '#[RE]^#=/RIY/' fails to match since it requires that R be followed by E. The next, and last, R-rule, '[R]=/R/', is the default; it matches any R not matched by earlier rules. Consequently, /R/ goes into the output string, and the scan moves past the R to A: R↑ATIO.

The search of the A-rules turns up no match before '[A]^+#=/EY/', which applies when A is followed by a single consonant, a front vowel (E, I, or Y), and another vowel. The program adds /EY/ to the output and moves the pointer past the A to T: RA↑TIO.

The first T-rule to match is '[TI]O=/SH/'. Consequently, /SH/ goes into the output and the pointer moves past TI to O: RATI↑O. The program does not search the I-rules, since the I occurs inside the brackets with the T; the string TI as a whole gets the pronunciation /SH/ and no output phonemes correspond to I alone.

The first match among the O-rules is '[O]#=/OW/'; the program outputs /OW/ and moves the pointer past the O to the blank at the end of the word: RATIO↑. The output string is /R/ /EY/ /SH/ /OW/, which represents the IPA /reʃo/, the correct transcription [12]. If the translation continued, the next matching rule would be in the set that passes blanks, commas, periods, and other punctuation into the output string as /⟨ ⟩/, /⟨,⟩/, /⟨.⟩ etc. The program would output /⟨ ⟩/ and move the pointer past the blank to the beginning of the next word, if any.

The IPA output string is input to a second pass that uses the same algorithm and rules of the same form to translate IPA to Votrax codes. The IPA-to-Votrax rules are fewer and more straightforward than the English-to-IPA rules (e.g., '[T]=[T]'). Since the synthesizer automatically varies the pronunciation of its "phonemes" to suit various contexts, the rules need not contain much context dependence. Some context-dependent rules have been included, however, to implement the manufacturer's suggestions about liquids, particularly L, adjacent to certain vowels. The complete set of rules may be found in [10].

Another program was used during rule development to insure that a rule change proposed to fix up a dozen mispronounced words would not ruin a hundred others previously translated correctly. This program, DICT, accepts a pattern like the left-hand side of a rule, but without brackets; it gives the same interpretations as TRANS to the same special symbols. After reading the pattern, DICT searches a file of words and outputs the words that contain a match.

The Brown Corpus comprises 500 samples of English text written in a wide variety of styles. Each sample is roughly 2000 words long, and the entire Corpus totals slightly more than a million words. The file we use lists the roughly 50 000 individual words occurring in the Corpus, arranged in decreasing order of frequency. The entry for each word contains some items of numerical information, including 1) *frequency*—the number of occurrences of the word in the Corpus, and

2) *number of texts*–the number of text samples, among the 500 comprising the Corpus, in which the word occurs.

One output that can be requested from TRANS is a "stat file"–a file listing every instance of every rule used in translating every word in a text file. A program STAT reads stat files and produces statistics on the relative importance of the rules. For each rule, STAT 1) counts the words in whose translation the rule was used, 2) sums the frequencies of those words, and 3) sums the number of text samples, among the five hundred in the Corpus, in which each of those words appears. The output comprises these three absolute results together with the relative results obtained by normalizing the absolute ones so that their sums over all rules are 1.

Listing of TRANS and DICT are available in [10]. Midway in the project, a Fasbol compiler [13] became available to us. We had found that, even in a program intended only as a research tool, execution speed was not a matter of complete indifference; we therefore partially rewrote our programs to take advantage of Fasbol; these versions are available from the authors. The change from interpretive execution to compilation has increased the programs' running speeds substantially–by a factor of 25 for TRANS. Translations now take a second or two per word, and there is no doubt that an implementation designed for efficiency rather than convenience of experimentation, by stripping away another layer of interpretive overhead, would have no trouble in surpassing real-time speech rates.

C. Rule Development

Our starting point, Version 1 of the rules, was a modification of Ainsworth's set. The main alterations were changes in the right-hand sides to Americanize the accent and additions to handle final S, ES, and ED correctly. Then began a development cycle with the following steps:

1) Translate: With Version 1, we translated the most frequent 4000 words in the Brown Corpus. With later versions, we included samples from deeper in the Corpus.

2) Examine Results: We had much of the translated output spoken by the synthesizer and listened to it, marking mistakes on a printed listing. Kenyon and Knott's pronouncing dictionary [12] was the arbiter in case of doubt or disagreement as to what constituted a "mistake." (The authors' linguistic backgrounds are diverse enough that disagreements were fairly frequent.) Later in the project, we grew proficient enough at reading the machine representation of IPA to risk checking some samples visually, but we never abandoned the practice of listening to at least part of the output from each version of the rules. The major goal was a good IPA transcription. In the few cases where a correct transcription still sounded strange, the IPA-to-Votrax rules were fixed up when possible, and the problem was otherwise blamed on the synthesizer.

3) Classify Errors: We divided the mispronounced words into lists with headings like "TH problem," "Silent E problem," "Long A problem," and "Stress problems." Then we scanned the lists to identify specific letter patterns being frequently mistranslated.

4) Modify: For a given frequently mistranslated letter pattern, we would find all sufficiently frequent words, mistranslated or not, that matched the pattern. If the correct pronunciations agreed in a majority of cases, or even a clear plurality, we wrote a new or altered rule to give that pronunciation; otherwise we tried a more specific context. For example, Version 1 had no rule for the EA combination, which has a great variety of pronunciations: *great, heart, ready, sea, earth*. Most words containing EA showed up on the "EA problem" list. We found the long-*e* pronunciation /i/ in roughly half of them. The addition of a rule '[EA]=/IY/' was justified, since it improved many words and did not harm the rest. *Meat* received the correct pronunciation /mit/, while *great* was no worse as /grit/ than it had been as /grεæt/. During the second round of development, many EA words still showed up as problems, but a search with DICT turned up the large number now getting the correct pronunciation. Looking for a more specific pattern, we found lots of EAD words on the problem list. A search of the Corpus for EAD words suggested adding a rule '[EA]D=/EH/', which fixes *ready*, changes one acceptable pronunciation of *lead* to another, and hurts a few previously correct words like *bead*.

In formulating new rules, we were helped by sources ranging from memories of elementary-school classes to scholarly studies such as [14]; see also [15]. Adding a new rule sometimes made it possible to delete an old one and frequently entailed altering or reordering old rules. The additions and alterations continued until the accumulation of changes made the interactions between rules hard to keep track of.

5) Iterate: Having produced a new version, we would start the cycle over by translating several thousand words. We went through the cycle twice, ending with Version 3. Before testing Version 3, we pruned the rules by looking at the Stat outputs for Version 2 and removing rules that were rarely used. Hence, rules for initial PT and initial X, although quite reliable, were thrown out for small importance.

D. Testing

We tested Version 3 by translating the 8000 most frequent words plus a thousand-word sample selected from the tail of the corpus–words with frequencies of 1 or 2 per million. The first 5000 words and the tail sample were scored like the translations by earlier versions: the criterion for correctness was a good IPA transcription, and, while we did not look up most words in a pronouncing dictionary, Kenyon and Knott [12] was the arbiter when questions arose. Numbers, symbols, and abbreviations were excluded from the scoring. Any transcription accepted by Kenyon and Knott was allowed, not just the preferred. Some deviations were allowed. The *horse:hoarse* distinction (/ɔr/ versus /or/) was ignored, as were the *Mary*: *merry*:*marry* distinction and similar distinctions involving vowels followed by R. Doubled consonants (/bItt.ɚ/instead of /bItɚ/for *bitter*) were not counted as errors. Otherwise we tried to be quite strict in scoring consonants and stressed vowels. Decisions about unstressed vowels were not so clear-cut, since the degrees of vowel reduction form a continuum. Translating an unstressed vowel with the full or stressed pronunciation seemed preferable to erroneously reducing a vowel to a schwa, since less loss of information was entailed. The unstressed vowels were scored accordingly. Instances of erroneous vowel reduction were marked as mistakes, but an unstressed vowel translated with the stressed pronunciation was often classed as a "stress problem", rather than a mistake, if vowel reduction upon destressing would give a good transcription. Thus, /æbaUt/ instead of /əbaUt/ for *about*, though marked as a stress problem, was not scored as an error. Some subjectivity entered here. Stress problems judged less severe than that in *about* were sometimes not marked at all; more severe ones were sometimes scored as errors.

For several of the samples we made phoneme-error counts in addition to the word-error counts. The criterion for "correctness" was the same as discussed above. For counting, a "phoneme" was any phoneme or phoneme-pair appearing in Table I.

III. Results

Table III gives the result of scoring IPA transcriptions of 1000-word samples from the Brown Corpus. The first three columns are based on a count of the number of distinct words correctly translated and the total number translated. The last three columns are based on the sums of the frequencies of the correctly translated words and of all the translated words. The frequencies were obtained from the Corpus; they give the number of times the word appeared and thus represent roughly parts per million. The first rows are based on successive 1000-word samples, starting from the beginning of the Corpus; the last is based on 1000 words selected from the tail of the Corpus (1/18 of the words with 2 occurrences per million and 1/36 of those with 1 per million).

Table IV is similar, but shows phoneme-error scores rather than word-error scores. The "number of phonemes scored" in the first column is the sum of the number of phoneme symbols marked correct and the number of errors counted. Since the errors include insertions and deletions as well as replacements, the sum is not strictly equal to either the number of phoneme symbols in the actual transcription or the number in a correct transcription; the difference, however, hardly matters.

Table V gives cumulative word-error scores; the last line is an estimate, derived from the foregoing, of the results that would have been obtained had the entire Corpus been translated and scored. The upper bounds were computed under the assumption that the error rate observed in the fifth 1000-word sample (see Table III) holds constant up to the beginning of the tail sample; the lower bounds assume that the error rate following the first 5000 words is equal to that observed in the tail. The figures 89 to 90 percent in the last column mean that, assuming the Corpus frequencies are representative, we would expect to correctly translate 89 to 90 percent of the words in a random sample of English text.

Table VI contains cumulative phoneme-error scores and includes whole-Corpus estimates similar to those in Table V.

Table VII gives results for the first 1000 words as translated at various stages of rule development.

Table VIII gives Version 3 of the English-to-IPA rules together with statistics for two samples: the first 8000 words of the Corpus and the 1000-word sample selected from the tail of the Corpus. For each sample, the columns marked "Abs." give the total number of distinct words that matched each rule and the sum of the frequencies of the words matching each rule. If a rule was used more than once in translating a word, that word contributed more than once to the word count and frequency sum for the given rule. Each "Relative" column represents the corresponding "Abs." column normalized to a total of 100 percent.

The IPA-to-Votrax rules, with similar statistics, are listed in [10].

IV. Discussion and Conclusions

Our results demonstrate that a very simple algorithm driven by a small set of letter-to-sound rules—fewer than 350—can produce correct IPA transcriptions of the great majority of English words without using a large pronouncing dictionary; with the same algorithm, driven by a smaller set of rules, the IPA transcription can be translated into a form acceptable to a commercial speech synthesizer. Of the thousand most frequent words in English, the process correctly pronounces more than 96 percent if words are counted according to their frequencies of occurrence.

Counting words according to their frequencies is one way of trying to give greater weight to more important words, as is clearly appropriate when the intended application is reading connected text rather than words selected at random from a dictionary. Simply weighting by word frequency does not take into account that the less frequent words tend to carry more information than the more frequent words, which include a large proportion of function words with low information content. This is not to say that frequency times information content, however defined, would always be a better weighting factor than frequency alone; mispronouncing the articles, for instance, would entail very little loss of information but would probably be intolerably distracting. For some purposes, nevertheless, phoneme-error rates may be a more appropriate measure than word-error rates, since the more informative words tend to be longer than the function words. We have therefore presented phoneme-error figures in addition to the word-error figures.

The word-error rate and the phoneme-error rate both rise with decreasing word frequency. However, since over two thirds of the words in a typical sample are among the most frequent thousand, the program's relatively poor performance on rare words does not drive the overall performance below about 90 percent of words, or about 97 percent of phonemes correct. Thus, on the average, the program mispronounces fewer than two words per sentence of ordinary written English. Most of the mispronunciations are single-phoneme errors and are easily correctable from context.

It has proved to be quite easy to modify the rules and experiment with different versions. As a result, we have been able, in passing from Version 1 to Version 3, to reduce the error rate for the thousand most frequent words from an initial 32 percent to the present 4 percent while increasing the number of rules by three quarters.

We were at first slightly disappointed and more than slightly puzzled by the discrepancy between our performance score of 68 percent for Version 1 of the rules (frequency-weighted score from Table V) and Ainsworth's reported scores of 89 to 92 percent for his set of rules [2]. Since our Version 1 is

TABLE III
WORD ERRORS: SCORES AND FREQUENCY-WEIGHTED SCORES FOR 1000-WORD SAMPLES OF BROWN CORPUS TRANSLATED BY VERSION 3 OF THE RULES

Sample	No. of Words Scored	No. of Words Correct	Fraction Correct	Total Freq. of Words Scored	Total Freq. of Correct Words	Fraction Correct (Freq.-Weighted)
1	976	847	86.8%	691,375	664,564	96.1%
2	974	808	83.0%	72,966	60,862	83.4%
3	973	744	76.5%	43,664	33,401	76.5%
4	988	757	76.6%	30,391	23,315	76.6%
5	971	707	72.8%	21,601	15,743	72.9%
Tail	922	599	65.0%	1,295	849	65.6%

TABLE IV
PHONEME ERRORS: SCORES AND FREQUENCY-WEIGHTED SCORES FOR 1000-WORD SAMPLES OF BROWN CORPUS TRANSLATED BY VERSION 3 OF THE RULES

Sample	No. of Phonemes Scored	No. of Phonemes Correct	Fraction Correct	Total Freq. of Phonemes Scored	Total Freq. of Correct Phonemes	Fraction Correct (Freq.-Weighted)
1	4,603	4,456	96.8%	2,096,545	2,066,811	98.6%
2	5,511	5,318	96.5%	411,288	397,108	96.6%
3	5,959	5,688	95.5%	266,692	254,539	95.4%
4	6,131	5,848	95.4%	188,017	179,370	95.4%
5	6,211	5,884	94.7%	138,375	131,146	94.8%
Tail	6,826	6,359	93.2%	9,499	8,855	93.2%

derived from and quite similar to Ainsworth's set, we had expected similar performance figures.

Three possible explanations suggest themselves. First, the difference between British and American pronunciation is more than a simple matter of dropping or retaining *r*'s and replacement of one sound by another. Ainsworth's rules, being adapted to British English, might, therefore, be in various subtle ways unamenable to Americanization by such straightforward changes as we made while setting up Version 1. Second, the question of what pronunciations of a word are acceptable is by no means cut and dried, even when one has a pronunciation dictionary at hand. Thus, although we had definite criteria in mind while scoring translations, we were not able to avoid subjectivity entirely. It is a dubious business at best to compare judgments of correctness arrived at independently under different circumstances by different judges having different expectations and different temperaments. Finally, the samples translated were different. The performance of a set of rules is sensitive to the vocabulary level of the material it is applied to; Table III illustrates this clearly. The Brown Corpus includes selections in a comprehensive range of styles, and Ainsworth's descriptions—"textbook on phonetics," "modern novel," "newspaper article on a political theme" [2]—do not pin down where in that range the sources of his samples fall; they may be written plainly, or their authors may have salted their language with rare words. The actual reason for the discrepancy in scores is probably some combination of these three explanations.

Further additions and refinements to the rules could reduce the error rate still further. After Version 3 of the rules was tested, however, it appeared that any improvement at all in stress, inflection, and rhythm would be more beneficial than reducing the error rate by a few more percent. The flat monotone produced by the system that has been described here is fatiguing to follow for long; some listeners have gone so far as to suggest that they would find any variation in stress and inflection, whether correct or not, more tolerable for extended listening.

There exist rules, though not yet a definitive set, for the placement of stress in English—see, for example [16]. However, the same goals of compactness and economy that led us to reject large on-line pronunciation dictionaries made us unwilling to deal with the lexical and syntactic information that would be necessary for a close approximation to correct English stress patterns. Even without near-perfect stress placement there seemed room for considerable improvement that might be achievable with a relatively Spartan stress-and-inflection scheme.

A few such schemes have been tested for their effects on listener preference and comprehension [17]. Both types of tests were conducted since naturalness and comprehensibility need not always go together, and one might be attained at the

TABLE V
WORD ERRORS: CUMULATIVE SCORES AND FREQUENCY-WEIGHTED SCORES FOR FIRST *n* THOUSAND WORDS OF BROWN CORPUS TRANSLATED BY VERSION 3 OF THE RULES

n	No. of Words Scored	No. of Words Correct	Fraction Correct	Total Freq. of Words Scored	Total Freq. of Correct Words	Fraction Correct (Freq.-Weighted)
1	976	847	86.8%	691,375	664,564	96.1%
2	1,950	1,655	84.9%	764,341	725,426	94.9%
3	2,923	2,399	82.1%	808,005	758,827	93.9%
4	3,911	3,156	80.7%	838,396	782,142	93.3%
5	4,882	3,863	79.1%	859,997	797,885	92.8%
Entire Corpus (Estimate)			66% to 69%			89% to 90%

TABLE VI
PHONEME ERRORS: CUMULATIVE SCORES AND FREQUENCY-WEIGHTED SCORES FOR FIRST *n* THOUSAND WORDS OF BROWN CORPUS TRANSLATED BY VERSION 3 OF THE RULES

n	No. of Phonemes Scored	No. of Phonemes Correct	Fraction Correct	Total Freq. of Phonemes Scored	Total Freq. of Correct Phonemes	Fraction Correct (Freq.-Weighted)
1	4,603	4,456	96.8%	2,096,545	2,066,811	98.6%
2	10,114	9,774	96.6%	2,507,833	2,463,919	98.2%
3	16,073	15,462	96.2%	2,774,525	2,718,458	98.0%
4	22,204	21,310	96.0%	2,962,542	2,897,828	97.8%
5	28,415	27,194	95.7%	3,100,917	3,028,974	97.7%
Entire Corpus (Estimate)			93% to 94%			96.6% to 96.9%

expense of the other. Six schemes were compared: 1) the unrelieved monotone; 2) alternating stress, where syllables were stressed and unstressed in strict alternation; 3) random stress, where stress was assigned randomly to syllables; 4) "stress algorithm," a term we applied to a combination of simple heuristics for English stress assignment; 5) hand-placed correct English stress; and 6) hand-placed English stress with additional timing adjustments to partially equalize the intervals between stressed syllables. For the comprehension tests, natural human speech was used in addition to speech synthesized according to these six schemes.

TABLE VII
WORD ERRORS: SCORES AND FREQUENCY-WEIGHTED SCORES FOR FIRST THOUSAND WORDS OF BROWN CORPUS TRANSLATED BY VARIOUS VERSIONS OF THE RULES

Version	No. of Rules*	No. of Words Scored	No. of Words Correct	Fraction Correct	Total Freq. of Words Scored	Total Freq. of Correct Words	Fraction Correct (Freq. - Weighted)
1	182	976	428	43.9%	691,375	470,575	68.1%
2	264	977	688	70.4%	691,497	606,287	87.7%
3	319	976	847	86.8%	691,375	664,564	96.1%

* These counts exclude rules for the 10 digits and for all punctuation symbols except .,-'? and blank.

The algorithm 4) bases stress assignment on two observations. The first is that a number of letter-to-sound rules, even in their present form, are good predictors for stressing and destressing. This is especially true of the rules with a schwa (/ə/) on the right-hand side, those for common function words, and those for common endings like ES and ED. The second is the tendency in English speech to stress approximately alternating syllables. Stress is assigned by using the rules where they apply and filling in stress markers on any remaining syllables by alternation. Unstressed syllables are given lower pitch and shorter duration than they would have if stressed. The timing of the stressed syllables is adjusted by further reduction of the durations of adjacent unstressed syllables and by lengthening any adjacent stressed syllables. Syllables before periods, question marks, semicolons, and colons are prolonged; those before question marks are given a rising inflection, and those before the other punctuation marks are given a falling inflection.

The listening-preference tests did not reveal a significant preference for alternating stress 2) or random stress 3) over the monotone 1), but indicated that the algorithm 4) and hand-placed English stress 5) were preferred over 1), 2), and 3). They did not show a significant preference between 4) and 5) but did show hand-placed English stress with timing adjustments 6) as significantly preferred over 4) and 5). The comprehension test results were consistent with these preference orderings but were not quite so clear-cut; those obtained after

TABLE VIII
STAT RESULTS: TWO SAMPLES FROM BROWN CORPUS TRANSLATED BY VERSION 3 OF THE RULES

*** ARULE ***								
[A] =/AX/	94	0.215	26051	0.907	30	0.461	41	0.455
[ARE] =/AA R/	1	0.002	4393	0.153	0	0.000	0	0.000
[AR]O=/AX R/	3	0.007	599	0.021	1	0.015	2	0.022
[AR]#=/EH R/	151	0.345	6320	0.220	10	0.154	15	0.166
^[AS]#=/EY S/	18	0.041	1334	0.046	0	0.000	0	0.000
[A]WA=/AX/	10	0.023	728	0.025	2	0.031	3	0.033
[AW]=/AO/	23	0.053	1256	0.044	6	0.092	10	0.111
:[ANY]=/EH N IY/	9	0.021	2954	0.103	0	0.000	0	0.000
[A]^+#=/EY/	221	0.505	8369	0.291	28	0.430	39	0.433
#:[ALLY]=/AX L IY/	46	0.105	1920	0.067	4	0.061	4	0.044
[AL]#=/AX L/	17	0.039	898	0.031	1	0.015	2	0.022
[AGAIN]=/AX G EH N/	2	0.005	1204	0.042	0	0.000	0	0.000
#:[AG]E=/IH JH/	49	0.112	1799	0.063	3	0.046	4	0.044
[A]^+:#=/AE/	193	0.441	7458	0.260	39	0.599	56	0.621
:[A]^+ =/EY/	89	0.204	9944	0.346	6	0.092	8	0.089
[A]^%=/EY/	232	0.530	8750	0.305	39	0.599	56	0.621
[ARR]=/AX R/	13	0.030	329	0.011	0	0.000	0	0.000
[ARR]=/AE R/	22	0.050	841	0.029	4	0.061	4	0.044
:[AR] =/AA R/	7	0.016	849	0.030	0	0.000	0	0.000
[AR] =/ER/	24	0.055	986	0.034	4	0.061	6	0.067
[AR]=/AA R/	211	0.482	10137	0.353	33	0.507	44	0.488
[AIR]=/EH R/	27	0.062	1244	0.043	5	0.077	7	0.078
[AI]=/EY/	163	0.373	6774	0.236	12	0.184	19	0.211
[AY]=/EY/	97	0.222	8739	0.304	12	0.184	18	0.200
[AU]=/AO/	59	0.135	2743	0.095	18	0.276	25	0.277
#:[AL] =/AX L/	201	0.460	11422	0.398	22	0.338	27	0.300
#:[ALS] =/AX L Z/	12	0.027	484	0.017	0	0.000	0	0.000
[ALK]=/AO K/	10	0.023	694	0.024	1	0.015	1	0.011
[AL]^=/AO L/	109	0.249	10348	0.360	24	0.369	32	0.355
:[ABLE]=/EY B AX L/	4	0.009	488	0.017	2	0.031	4	0.044
[ABLE]=/AX B AX L/	45	0.103	1342	0.047	4	0.061	5	0.055
[ANG]+=/EY N JH/	29	0.066	1495	0.052	1	0.015	2	0.022
[A]=/AE/	1482	3.389	118519	4.125	263	4.039	366	4.060
	3673	8.398	261411	9.097	574	8.815	800	8.874
*** BRULE ***								
[BE]^#=/B IH/	35	0.080	4727	0.165	5	0.077	7	0.078
[BEING]=/B IY IH NX/	2	0.005	748	0.026	0	0.000	0	0.000
[BOTH] =/B OW TH/	1	0.002	730	0.025	0	0.000	0	0.000
[BUS]#=/B IH Z/	4	0.009	484	0.017	0	0.000	0	0.000
[BUIL]=/B IH L/	6	0.014	481	0.017	1	0.015	1	0.011
[B]=/B/	729	1.667	50010	1.740	146	2.242	207	2.296
	777	1.777	57180	1.990	152	2.334	215	2.385
*** CRULE ***								
[CH]^=/K/	9	0.021	392	0.014	1	0.015	2	0.022
^E[CH]=/K/	10	0.023	451	0.016	2	0.031	3	0.033
[CH]=/CH/	215	0.492	16131	0.561	38	0.584	57	0.632
S[CI]#=/S AY/	5	0.011	305	0.011	0	0.000	0	0.000
[CI]A=/SH/	35	0.080	1763	0.061	5	0.077	7	0.078
[CI]O=/SH/	10	0.023	230	0.008	2	0.031	4	0.044
[CI]EN=/SH/	7	0.016	307	0.011	1	0.015	1	0.011
[C]+=/S/	475	1.086	23550	0.820	47	0.722	66	0.732
[CK]=/K/	98	0.224	4217	0.147	33	0.507	46	0.510
[COM]%=/K AH M/	13	0.030	1706	0.059	0	0.000	0	0.000
[C]=/K/	1482	3.389	65195	2.269	174	2.672	251	2.784
	2359	5.394	114247	3.976	303	4.653	437	4.847
*** DRULE ***								
#:[DED] =/D IH D/	51	0.117	1927	0.067	8	0.123	12	0.133
.E[D] =/D/	312	0.713	12985	0.452	37	0.568	55	0.610
#^:E[D] =/T/	140	0.320	6040	0.210	12	0.184	18	0.200
[DE]^#=/D IH/	124	0.284	4867	0.169	13	0.200	19	0.211
[DO] =/D UW/	1	0.002	1363	0.047	0	0.000	0	0.000
[DOES]=/D AH Z/	2	0.005	572	0.020	0	0.000	0	0.000
[DOING]=/D UW IH NX/	1	0.002	163	0.006	0	0.000	0	0.000
[DOW]=/D AW/	4	0.009	964	0.034	0	0.000	0	0.000
[DU]A=/JH UW/	12	0.027	503	0.018	0	0.000	0	0.000
[D]=/D/	1301	2.975	102440	3.565	201	3.087	275	3.050
	1948	4.454	131824	4.588	271	4.162	379	4.204

TABLE VIII (*Continued*)

*** ERULE ***								
#:[E] =/ /	1006	2.300	73857	2.570	108	1.658	149	1.653
' ^:[E] =/ /	7	0.016	519	0.018	0	0.000	0	0.000
:[E] =/IY/	19	0.043	23483	0.817	4	0.061	6	0.067
#[ED] =/D/	14	0.032	641	0.022	2	0.031	3	0.033
#:[E]D =/ /	446	1.020	18502	0.644	45	0.691	68	0.754
[EV]ER=/EH V/	20	0.046	3258	0.113	2	0.031	3	0.033
[E]^%=/IY/	106	0.242	4302	0.150	17	0.261	30	0.333
[ERI]#=/IY R IY/	21	0.048	1508	0.052	0	0.000	0	0.000
[ERI]=/EH R IH/	24	0.055	1423	0.050	4	0.061	4	0.044
#:[ER]#=/ER/	115	0.263	6410	0.223	11	0.169	15	0.166
[ER]#=/EH R/	17	0.039	1110	0.039	2	0.031	4	0.044
[ER]=/ER/	622	1.422	33594	1.169	105	1.612	153	1.697
[EVEN]=/IY V EH N/	7	0.016	1564	0.054	0	0.000	0	0.000
#:[E]W=/ /	10	0.023	173	0.006	1	0.015	1	0.011
@[EW]=/UW/	20	0.046	2819	0.098	6	0.092	10	0.111
[EW]=/Y UW/	1	0.002	601	0.021	0	0.000	0	0.000
[E]O=/IY/	27	0.062	792	0.028	6	0.092	9	0.100
#:&[ES] =/IH Z/	116	0.265	4265	0.148	11	0.169	15	0.166
#:[E]S =/ /	264	0.604	11065	0.385	38	0.584	53	0.588
#:[ELY] =/L IY/	45	0.103	1834	0.064	5	0.077	8	0.089
#:[EMENT]=/M EH N T/	37	0.085	1437	0.050	2	0.031	3	0.033
[EFUL]=/F UH L/	7	0.016	281	0.010	1	0.015	2	0.022
[EE]=/IY/	168	0.384	13544	0.471	22	0.338	29	0.322
[EARN]=/ER N/	8	0.018	345	0.012	0	0.000	0	0.000
[EAR]^=/ER/	7	0.016	751	0.026	0	0.000	0	0.000
[EAD]=/EH D/	29	0.066	2297	0.080	4	0.061	7	0.078
#:[EA] =/IY AX/	3	0.007	530	0.018	0	0.000	0	0.000
[EA]SU=/EH/	9	0.021	440	0.015	0	0.000	0	0.000
[EA]=/IY/	302	0.691	17378	0.605	36	0.553	46	0.510
[EIGH]=/EY/	16	0.037	534	0.019	1	0.015	2	0.022
[EI]=/IY/	31	0.071	1349	0.047	5	0.077	7	0.078
[EYE]=/AY/	3	0.007	533	0.019	1	0.015	1	0.011
[EY]=/IY/	30	0.069	1169	0.041	15	0.230	18	0.200
[EU]=/Y UW/	11	0.025	364	0.013	7	0.107	10	0.111
[E]=/EH/	2065	4.722	95200	3.313	301	4.622	403	4.470
	5633	12.880	327872	11.410	762	11.701	1059	11.747
*** FRULE ***								
[FUL]=/F UH L/	29	0.066	1043	0.036	4	0.061	7	0.078
[F]=/F/	736	1.683	58778	2.046	115	1.766	159	1.764
	765	1.749	59821	2.082	119	1.827	166	1.841
*** GRULE ***								
[GIV]=/G IH V/	6	0.014	1015	0.035	0	0.000	0	0.000
[G]I^=/G/	8	0.018	475	0.017	3	0.046	4	0.044
[GE]T=/G EH/	12	0.027	1504	0.052	0	0.000	0	0.000
SU[GGES]=/G JH EH S/	6	0.014	258	0.009	0	0.000	0	0.000
[GG]=/G/	20	0.046	399	0.014	4	0.061	5	0.055
B#[G]=/G/	10	0.023	1102	0.038	1	0.015	1	0.011
[G]+=/JH/	176	0.402	7355	0.256	34	0.522	49	0.544
[GREAT]=/G R EY T/	5	0.011	1014	0.035	0	0.000	0	0.000
#[GH]=/ /	11	0.025	522	0.018	1	0.015	1	0.011
[G]=/G/	347	0.793	15701	0.546	73	1.121	101	1.120
	601	1.374	29345	1.021	116	1.781	161	1.786
*** HRULE ***								
[HAV]=/HH AE V/	5	0.011	4284	0.149	0	0.000	0	0.000
[HERE]=/HH IY R/	2	0.005	761	0.026	0	0.000	0	0.000
[HOUR]=/AW ER/	2	0.005	319	0.011	1	0.015	2	0.022
[HOW]=/HH AW/	8	0.018	1583	0.055	1	0.015	2	0.022
[H]#=/HH/	296	0.677	45711	1.591	58	0.891	76	0.843
[H]=/ /	21	0.048	976	0.034	10	0.154	12	0.133
	334	0.764	53634	1.867	70	1.075	92	1.021

TABLE VIII (*Continued*)

*** IRULE ***								
[IN]=/IH N/	202	0.462	31259	1.088	27	0.415	36	0.399
[I] =/AY/	6	0.014	5894	0.205	2	0.031	3	0.033
[IN]D=/AY N/	22	0.050	2022	0.070	4	0.061	6	0.067
[IER]=/IY ER/	11	0.025	419	0.015	6	0.092	9	0.100
#:R[IED] =/IY D/	6	0.014	348	0.012	1	0.015	1	0.011
[IED] =/AY D/	24	0.055	1009	0.035	7	0.107	7	0.078
[IEN]=/IY EH N/	17	0.039	700	0.024	1	0.015	1	0.011
[IE]T=/AY EH/	13	0.030	779	0.027	2	0.031	3	0.033
:[I]%=/AY/	10	0.023	277	0.010	1	0.015	1	0.011
[I]%=/IY/	88	0.201	2808	0.098	17	0.261	27	0.300
[IE]=/IY/	36	0.082	1811	0.063	11	0.169	16	0.1.77
[I]^+:#=/IH/	384	0.878	15196	0.529	56	0.860	71	0.788
[IR]#=/AY R/	51	0.117	2006	0.070	4	0.061	6	0.067
[IZ]%=/AY Z/	19	0.043	697	0.024	7	0.107	9	0.100
[IS]%=/AY Z/	32	0.073	1027	0.036	4	0.061	5	0.055
[I]D%=/AY/	40	0.091	2544	0.089	4	0.061	7	0.078
+^[I]^+=/IH/	74	0.169	2855	0.099	11	0.169	16	0.1.77
[I]T%=/AY/	24	0.055	2043	0.071	6	0.092	7	0.078
#^:[I]^+=/IH/	232	0.530	9645	0.336	20	0.307	29	0.322
[I]^+=/AY/	116	0.265	10713	0.373	16	0.246	25	0.277
[IR]=/ER/	42	0.096	3221	0.112	12	0.184	18	0.200
[IGH]=/AY/	55	0.126	4271	0.149	6	0.092	8	0.089
[ILD]=/AY L D/	11	0.025	810	0.028	0	0.000	0	0.000
[IGN] =/AY N/	3	0.007	226	0.008	0	0.000	0	0.000
[IGN]^=/AY N/	4	0.009	176	0.006	0	0.000	0	0.000
[IGN]%=/AY N/	4	0.009	216	0.008	0	0.000	0	0.000
[IQUE]=/IY K/	4	0.009	229	0.008	3	0.046	6	0.067
[I]=/IH/	2038	4.660	128923	4.487	356	5.467	493	5.469
	3568	8.158	232124	8.078	584	8.968	810	8.985
*** JRULE ***								
[J]=/JH/	125	0.286	6066	0.211	20	0.307	28	0.3.11
	125	0.286	6066	0.211	20	0.307	28	0.311
*** KRULE ***								
[K]N=/ /	13	0.030	1847	0.064	2	0.031	3	0.033
[K]=/K/	224	0.512	13401	0.466	62	0.952	81	0.899
	237	0.542	15248	0.531	64	0.983	84	0.932
*** LRULE ***								
[LO]C#=/L OW/	9	0.021	514	0.018	0	0.000	0	0.000
L[L]=/ /	236	0.540	17526	0.610	38	0.584	51	0.566
#^:[L]%=/AX L/	108	0.247	6084	0.212	21	0.322	26	0.288
[LEAD]=/L IY D/	7	0.016	515	0.018	2	0.031	2	0.022
[L]=/L/	1755	4.013	85646	2.981	297	4.561	422	4.681
	2115	4.836	110285	3.838	358	5.498	501	5.557
*** MRULE ***								
[MOV]=/M UW V/	12	0.027	930	0.032	0	0.000	0	0.000
[M]=/M/	1370	3.133	88465	3.079	231	3.547	317	3.516
	1382	3.160	89395	3.111	231	3.547	317	3.516
*** NRULE ***								
E[NG]+=/N JH/	9	0.021	270	0.009	0	0.000	0	0.000
[NG]R=/NX G/	9	0.021	353	0.012	2	0.031	3	0.033
[NG]#=/NX G/	30	0.069	1036	0.036	6	0.092	8	0.089
[NGL]%=/NX G AX L/	4	0.009	254	0.009	2	0.031	3	0.033
[NG]=/NX/	526	1.203	23241	0.809	84	1.290	113	1.253
[NK]=/NX K/	38	0.087	1577	0.055	8	0.123	11	0.122
[NOW] =/N AW/	1	0.002	1314	0.046	0	0.000	0	0.000
[N]=/N/	2446	5.593	170584	5.937	359	5.513	490	5.435
	3063	7.004	198629	6.913	461	7.079	628	6.966

TABLE VIII (*Continued*)

*** ORULE ***								
[OF] =/AX V/	2	0.005	36427	1.268	2	0.031	3	0.033
[OROUGH]=/ER OW/	2	0.005	61	0.002	0	0.000	0	0.000
#:[OR] =/ER/	69	0.158	2711	0.094	4	0.061	5	0.055
#:[ORS] =/ER Z/	22	0.050	624	0.022	3	0.046	5	0.055
[OR]=/AO R/	360	0.823	32460	1.130	55	0.845	87	0.965
[ONE]=/W AH N/	4	0.009	3487	0.121	0	0.000	0	0.000
[OW]=/OW/	112	0.256	7450	0.259	25	0.384	31	0.344
[OVER]=/OW V ER/	9	0.021	1398	0.049	4	0.061	5	0.055
[OV]=/AH V/	70	0.160	3713	0.129	11	0.169	17	0.189
[O]^%=/OW/	134	0.306	7003	0.244	11	0.169	13	0.144
[O]^EN=/OW/	32	0.073	1849	0.064	5	0.077	7	0.078
[O]^I#=/OW/	40	0.091	1728	0.060	12	0.184	19	0.211
[OL]D=/OW L/	27	0.062	2161	0.075	2	0.031	2	0.022
[OUGHT]=/AO T/	9	0.021	1072	0.037	0	0.000	0	0.000
[OUGH]=/AH F/	5	0.011	544	0.019	1	0.015	2	0.022
[OU]=/AW/	15	0.034	3895	0.136	3	0.046	5	0.055
H[OU]S#=/AW/	8	0.018	932	0.032	4	0.061	4	0.044
[OUS]=/AX S/	56	0.128	2031	0.071	8	0.123	11	0.122
[OUR]=/AO R/	28	0.064	1955	0.068	3	0.046	5	0.055
[OULD]=/UH D/	9	0.021	5649	0.197	0	0.000	0	0.000
[OU]^L=/AH/	10	0.023	443	0.015	1	0.015	1	0.011
[OUP]=/UW P/	3	0.007	531	0.018	1	0.015	1	0.011
[OU]=/AW/	107	0.245	8077	0.281	16	0.246	23	0.255
[OY]=/OY/	28	0.064	1137	0.040	3	0.046	4	0.044
[OING]=/OW IH NX/	3	0.007	422	0.015	0	0.000	0	0.000
[OI]=/OY/	42	0.096	1903	0.066	9	0.138	10	0.111
[OOR]=/AO R/	12	0.027	745	0.026	2	0.031	3	0.033
[OOK]=/UH K/	13	0.030	1948	0.068	3	0.046	4	0.044
[OOD]=/UH D/	19	0.043	1847	0.064	3	0.046	5	0.055
[OO]=/UW/	60	0.137	3764	0.131	16	0.246	18	0.200
[O]E=/OW/	20	0.046	772	0.027	1	0.015	1	0.011
[O] =/OW/	49	0.112	7433	0.259	32	0.491	44	0.488
[OA]=/OW/	47	0.107	1964	0.068	8	0.123	9	0.100
[ONLY]=/OW N L IY/	1	0.002	1747	0.061	0	0.000	0	0.000
[ONCE]=/W AH N S/	1	0.002	499	0.017	1	0.015	1	0.011
[ON ' T]=/OW N T/	2	0.005	594	0.021	0	0.000	0	0.000
C[O]N=/AA/	179	0.409	7030	0.245	17	0.261	24	0.266
[O]NG=/AO/	22	0.050	2475	0.086	3	0.046	4	0.044
^:[O]N=/AH/	57	0.130	2364	0.082	14	0.215	21	0.233
I[ON]=/AX N/	362	0.828	14961	0.521	25	0.384	33	0.366
#:[ON] =/AX N/	70	0.160	2648	0.092	19	0.292	26	0.288
#^[ON]=/AX N/	23	0.053	691	0.024	10	0.154	14	0.155
[O]ST =/OW/	8	0.018	2137	0.074	1	0.015	1	0.011
[OF]^=/AO F/	17	0.039	2065	0.072	2	0.031	2	0.022
[OTHER]=/AH DH ER/	12	0.027	3231	0.112	1	0.015	1	0.011
[OSS] =/AO S/	6	0.014	520	0.018	0	0.000	0	0.000
#^:[OM]=/AH M/	49	0.112	1627	0.057	8	0.123	13	0.144
[O]=/AA/	850	1.944	51239	1.783	122	1.873	165	1.830
	3085	7.054	241964	8.421	471	7.233	649	7.199
*** PRULE ***								
[PH]=/F/	59	0.135	1717	0.060	21	0.322	29	0.322
[PEOP]=/P IY P/	3	0.007	902	0.031	1	0.015	1	0.011
[POW]=/P AW/	6	0.014	535	0.019	0	0.000	0	0.000
[PUT] =/P UH T/	3	0.007	492	0.017	0	0.000	0	0.000
[P]=/P/	1556	3.558	69000	2.401	194	2.979	269	2.984
	1627	3.720	72646	2.528	216	3.317	299	3.317
*** QRULE ***								
[QUAR]=/K W AO R/	7	0.016	314	0.011	0	0.000	0	0.000
[QU]=/K W/	76	0.174	3287	0.114	10	0.154	14	0.155
[Q]=/K/	2	0.005	35	0.001	1	0.015	2	0.022
	85	0.194	3636	0.127	11	0.169	16	0.177
*** RRULE ***								
[RE]^#=/R IY/	186	0.425	8287	0.288	14	0.215	19	0.211
[R]=/R/	1497	3.423	73680	2.564	228	3.501	315	3.494
	1683	3.848	81967	2.853	242	3.716	334	3.705

TABLE VIII (*Continued*)

*** SRULE ***								
[SH]=/SH/	177	0.405	10754	0.374	25	0.384	31	0.344
#[SION]=/ZH AX N/	23	0.053	972	0.034	0	0.000	0	0.000
[SOME]=/S AH M/	12	0.027	2772	0.096	2	0.031	4	0.044
#[SUR]#=/ZH ER/	11	0.025	476	0.017	0	0.000	0	0.000
[SUR]#=/SH ER/	10	0.023	709	0.025	1	0.015	2	0.022
#[SU]#=/ZH UW/	5	0.011	416	0.014	0	0.000	0	0.000
#[SSU]#=/SH UW/	5	0.011	322	0.011	0	0.000	0	0.000
#[SED] =/Z D/	26	0.059	1686	0.059	3	0.046	4	0.044
#[S]#=/Z/	271	0.620	13840	0.482	43	0.660	59	0.654
[SAID]=/S EH D/	1	0.002	1961	0.068	0	0.000	0	0.000
^[SION]=/SH AX N/	43	0.098	1415	0.049	5	0.077	7	0.078
[S]S=/ /	248	0.567	10255	0.357	33	0.507	42	0.466
.[S] =/Z/	512	1.171	21193	0.738	63	0.967	85	0.943
#:.E[S] =/Z/	138	0.316	5887	0.205	16	0.246	22	0.244
#^:##[S] =/Z/	107	0.245	4437	0.154	20	0.307	33	0.366
#^:#[S] =/S/	89	0.204	3773	0.131	19	0.292	25	0.277
U[S] =/S/	3	0.007	778	0.027	1	0.015	2	0.022
:#[S] =/Z/	39	0.089	38870	1.353	8	0.123	10	0.111
[SCH]=/S K/	9	0.021	883	0.031	2	0.031	3	0.033
[S]C+=/ /	20	0.046	723	0.025	4	0.061	6	0.067
#[SM]=/Z M/	26	0.059	514	0.018	7	0.107	11	0.122
#[SN] '=/Z AX N/	3	0.007	271	0.009	0	0.000	0	0.000
[S]=/S/	2063	4.717	104475	3.636	307	4.714	418	4.637
	3841	8.782	227382	7.913	559	8.584	764	8.475
*** TRULE ***								
[THE] =/DH AX/	1	0.002	69971	2.435	2	0.031	3	0.033
[TO] =/T UW/	14	0.032	28177	0.981	2	0.031	3	0.033
[THAT] =/DH AE T/	2	0.005	10781	0.375	0	0.000	0	0.000
[THIS] =/DH IH S/	1	0.002	5146	0.179	0	0.000	0	0.000
[THEY]=/DH EY/	5	0.011	3761	0.131	0	0.000	0	0.000
[THERE]=/DH EH R/	8	0.018	3142	0.109	0	0.000	0	0.000
[THER]=/DH ER/	27	0.062	2408	0.084	3	0.046	5	0.055
[THEIR]=/DH EH R/	2	0.005	2691	0.094	0	0.000	0	0.000
[THAN] =/DH AE N/	1	0.002	1789	0.062	1	0.015	2	0.022
[THEM] =/DH EH M/	1	0.002	1789	0.062	0	0.000	0	0.000
[THESE] =/DH IY Z/	1	0.002	1573	0.055	0	0.000	0	0.000
[THEN]=/DH EH N/	1	0.002	1377	0.048	0	0.000	0	0.000
[THROUGH]=/TH R UW/	2	0.005	1110	0.039	0	0.000	0	0.000
[THOSE]=/DH OW Z/	1	0.002	850	0.030	0	0.000	0	0.000
[THOUGH] =/DH OW/	2	0.005	761	0.026	0	0.000	0	0.000
[THUS]=/DH AH S/	1	0.002	312	0.011	0	0.000	0	0.000
[TH]=/TH/	191	0.437	19586	0.682	28	0.430	37	0.410
#:[TED] =/T IH D/	186	0.425	6418	0.223	11	0.169	17	0.189
S[TI]#N=/CH/	12	0.027	756	0.026	1	0.015	1	0.011
[TI]O=/SH/	338	0.773	13438	0.468	20	0.307	28	0.311
[TI]A=/SH/	17	0.039	603	0.021	2	0.031	2	0.022
[TIEN]=/SH AX N/	4	0.009	165	0.006	0	0.000	0	0.000
[TUR]#=/CH ER/	55	0.126	2573	0.090	3	0.046	6	0.067
[TU]A=/CH UW/	15	0.034	858	0.030	4	0.061	6	0.067
[TWO]=/T UW/	2	0.005	1424	0.050	2	0.031	2	0.022
[T]=/T/	3064	7.006	183179	6.375	406	6.235	555	6.156
	3954	9.041	364638	12.690	485	7.448	667	7.399
*** URULE ***								
[UN]I=/Y UW N/	15	0.034	1461	0.051	2	0.031	3	0.033
[UN]=/AH N/	49	0.112	2462	0.086	17	0.261	23	0.255
[UPON]=/AX P AO N/	1	0.002	495	0.017	0	0.000	0	0.000
@[UR]#=/UH R/	15	0.034	1084	0.038	4	0.061	5	0.055
[UR]#=/Y UH R/	26	0.059	980	0.034	2	0.031	3	0.033
[UR]=/ER/	109	0.249	4572	0.159	17	0.261	23	0.255
[U]^ =/AH/	70	0.160	9270	0.323	13	0.200	16	0.177
[U]^^=/AH/	366	0.837	17715	0.617	59	0.906	84	0.932
[UY]=/AY/	5	0.011	182	0.006	2	0.031	3	0.033
G[U]#=/ /	16	0.037	470	0.016	1	0.015	1	0.011
G[U]%=/ /	11	0.025	270	0.009	0	0.000	0	0.000
G[U]#=/W/	9	0.021	278	0.010	2	0.031	2	0.022
#N[U]=/Y UW/	25	0.057	1149	0.040	3	0.046	3	0.033
@[U]=/UW/	198	0.453	7998	0.278	23	0.353	31	0.344
[U]=/Y UW/	149	0.341	7024	0.244	19	0.292	27	0.300
	1064	2.433	55410	1.928	164	2.518	224	2.485
*** VRULE ***								
[VIEW]=/V Y UW/	9	0.021	411	0.014	0	0.000	0	0.000
[V]=/V/	550	1.258	22264	0.775	66	1.014	91	1.009
	559	1.278	22675	0.789	66	1.014	91	1.009

TABLE VIII (*Continued*)

*** WRULE ***								
[WERE]=/W ER/	2	0.005	3306	0.115	0	0.000	0	0.000
[WA]S=/W AA/	8	0.018	10343	0.360	1	0.015	1	0.011
[WA]T=/W AA/	8	0.018	794	0.028	2	0.031	3	0.033
[WHERE]=/WH EH R/	8	0.018	1216	0.042	0	0.000	0	0.000
[WHAT]=/WH AA T/	4	0.009	2200	0.077	1	0.015	1	0.011
[WHOL]=/HH OW L/	3	0.007	344	0.012	1	0.015	1	0.011
[WHO]=/HH UW/	5	0.011	2681	0.093	0	0.000	0	0.000
[WH]=/WH/	18	0.041	7925	0.276	5	0.077	7	0.078
[WAR]=/W AO R/	31	0.071	1372	0.048	2	0.031	3	0.033
[WOR]^=/W ER/	25	0.057	3136	0.109	7	0.107	8	0.089
[WR]=/R/	12	0.027	961	0.033	1	0.015	1	0.011
[W]=/W/	222	0.508	29047	1.011	41	0.630	56	0.621
	346	0.791	63325	2.204	61	0.937	81	0.899
*** XRULE ***								
[X]=/K S/	179	0.409	7242	0.252	19	0.292	26	0.288
	179	0.409	7242	0.252	19	0.292	26	0.288
*** YRULE ***								
[YOUNG]=/Y AH NX/	4	0.009	461	0.016	0	0.000	0	0.000
[YOU]=/Y UW/	11	0.025	4749	0.165	0	0.000	0	0.000
[YES]=/Y EH S/	2	0.005	227	0.008	0	0.000	0	0.000
[Y]=/Y/	22	0.050	2764	0.096	4	0.061	6	0.067
#^:[Y] =/IY/	514	1.175	24405	0.849	67	1.029	97	1.076
#^:[Y]I=/IY/	8	0.018	245	0.009	3	0.046	3	0.033
:[Y] =/AY/	10	0.023	7400	0.258	2	0.031	2	0.022
:[Y]#=/AY/	8	0.018	347	0.012	2	0.031	3	0.033
:[Y]^+:#=/IH/	8	0.018	304	0.011	4	0.061	6	0.067
:[Y]^#=/AY/	24	0.055	930	0.032	5	0.077	6	0.067
[Y]=/IH/	63	0.144	2235	0.078	21	0.322	30	0.333
	674	1.541	44067	1.534	108	1.658	153	1.697
*** ZRULE ***								
[Z]=/Z/	58	0.133	1419	0.049	25	0.384	34	0.377
	58	0.133	1419	0.049	25	0.384	34	0.377
	=======	======	=======	======	=======	======	=======	======
	43735	100.	2873452	100.	6512	100.	9015	100.

the subjects had had a brief opportunity to practice listening were not statistically significant.

We have drawn three conclusions: 1) Not only stress and pitch, but rhythm and timing are important to producing acceptable synthetic speech. 2) Not just any stress pattern will do; for instance neither random stress nor alternating stress was significantly better than the monotone. 3) A simple algorithm for adjusting stress, inflection, and rhythm can significantly improve not only the listener acceptability of synthesized speech, but also, at least for naive listeners, its comprehensibility. The reader is referred to [17] for a full discussion.

A list of words pertinent to topics of interest to the Navy has been assembled, and the relative frequencies of the words have been estimated [18]. One of the possible applications for such a list would be in tailoring a version of the letter-to-sound rules for special applications within the Navy. While there is no reason to expect that the statistics of the ordinary words in "Naval English" would require much reworking of the rules, it is quite certain that acronyms would take special treatment. The pronunciations people give to "unpronounceable" combinations like WWMCCS (/wɪmɪks/) are too arbitrary for any systematic procedure to have much hope of duplicating them. A more reasonable goal is to pronounce pronounceable combinations plausibly and spell out unpronounceable ones. One simple expedient is to pronounce each consonant as its name when the context is an isolated cluster consisting entirely of consonants. This already catches a good number of important acronyms and abbreviations (e.g., NRL!), and the idea could be pushed further.

Acknowledgment

The authors are grateful to D. Stryker for writing the software for the TI-960A that controls the Speech-Synthesis Lab System. They also thank Dr. D. Woods for his advice on Americanizing Ainsworth's rules and Dr. I. Stockman for her help in setting up the IPA-to-Votrax correspondence. A program that tabulates scores for translated words was contributed by Ms. K. L. Babington; we are also indebted to her for help with the scoring.

References

[1] G. S. Kang, "Application of linear predictive encoding to a narrowband voice digitizer," Naval Res. Lab., Washington, DC, NRL Rep. 7774, Oct. 1974.

[2] W. A. Ainsworth, "A system for converting English text into speech," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 288-290, June 1974.

[3] M. D. McIlroy, "Synthetic English speech by rule," Bell Tele. Lab., Inc., Murray Hill, NJ, Mar. 1974.

[4] F. F. Lee, "Machine to man communication by speech Part I: Generation of segmental phonemes from text," in *1968 Spring Joint Computer Conf.*, pp. 333-338.

[5] J. Allen, "Machine to man communication by speech, Part II: Synthesis of prosodic features of speech by rule," in *1968 Spring Joint Computer Conf.*, pp. 339-344.

[6] F. F. Lee, "Reading machine: From text to speech," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 275-282, Dec. 1969.

[7] J. Allen, "Speech synthesis from unrestricted text," *IEEE Convention Digest*, 1971, pp. 108-109.

[8] —, "Reading machines for the blind: The technical problems and the methods adopted for their solution," *IEEE Trans. Audio Electroacoust.*, vol. AU-21, pp. 259-264, June 1973.

[9] H. Kucera and W. N. Francis, *Computational Analysis of Present-Day American English.* Providence, RI: Brown Univ. Press, 1967.

[10] H. Elovitz, R. Johnson, A. McHugh, and J. Shore, "Automatic translation of English text to phonetics by means of letter-to-sound rules," Naval Res. Lab., Washington, DC, NRL Rep. 7948, Jan. 1976.

[11] W. A. Ainsworth, private communication, Apr. 1974.

[12] J. S. Kenyon and T. A. Knott, *A Pronouncing Dictionary of American English.* Springfield, MA: Merriam, 1951.

[13] P. J. Santos, Jr., "FASBOL II, a SNOBOL compiler for the PDP-10," Digital Equipment Computer Users' Society, DECUS 10-179, Dec. 1972.

[14] R. Venezky, "A study of English spelling-to-sound correspondence on historical principles," Stanford Univ., Stanford, CA, 1965.

[15] —, *The Structure of English Orthography.* The Hague, The Netherlands: Mouton, 1970.

[16] N. Chomsky and M. Halle, *The Sound Patterns of English.* New York: Harper & Row, 1968.

[17] A McHugh, "Listener preference and comprehension tests of stress algorithms for a text-to-phonetic speech synthesis program," Naval Res. Lab., Washington, DC, NRL Rep. 8015, Sept. 1976.

[18] —, "Frequency ranked and alphabetical lists of words judged by navy officers to be frequent in navy usage," to be published.

Quantization and Bit Allocation in Speech Processing

AUGUSTINE H. GRAY, JR., MEMBER, IEEE, AND JOHN D. MARKEL, MEMBER, IEEE

***Abstract*—The topic of quantization and bit allocation in speech processing is studied using an L_2 norm. Closed-form expressions are derived for the root mean square (rms) spectral deviation due to variations in one, two, or multiple parameters. For one-parameter variation, the reflection coefficients, log area ratios, and inverse sine coefficients are studied. It is shown that, depending upon the criterion chosen, either log area ratios or inverse sine quantization can be viewed as optimal. From a practical point of view, it is shown experimentally that very little difference exists among the various quantization methods beyond the second coefficient.**

Two-parameter variations are studied in terms of formant frequency and bandwidth movement and in terms of a two-pair quantization scheme. A lower bound on the number of quantization levels required to satisfy a given maximum spectral deviation is derived along with the two-pair quantization scheme which approximately satisfies the bound. It is shown theoretically that the two-pair quantization scheme has a 10-bit superiority over other above-mentioned quantization schemes in the sense of theoretically assuring that a maximum overall log spectral deviation will not be exceeded.

I. Introduction

THE QUANTIZATION properties of transmission parameters in linear prediction speech compression systems have been discussed in a recent paper by Viswanathan and Makhoul [1] on the basis of a spectral sensitivity analysis using an L_1 norm on log spectral differences. The emphasis was on sensitivity to the reflection coefficient parameter set, $\{k_1, k_2, \cdots, k_M\}$, which can define an all-pole filter model. The reflection coefficients have the desirable property of retaining stability under quantization. Spectral sensitivity curves were numerically obtained and used to show that log area quantization of the reflection coefficients was optimal in the sense of minimizing a maximum spectral error over the entire range of possible values for the reflection coefficients, -1 to $+1$. In addition, based upon the assumption that total spectral deviation due to changes in all the parameters can be approximated by the sum of individual deviations, a bit-allocation scheme was proposed for minimizing the total spectral deviation.

The topic of quantization and bit allocation is also studied here, but using an L_2 as opposed to the L_1 norm. This seemingly minor change has a significant effect in that tractable

Manuscript received January 12, 1976; revised June 7, 1976. This work was supported by the Department of Defense under Contract MDA904-76-C-0259.

A. H. Gray, Jr. is with the Department of Electrical Engineering, University of California, Santa Barbara, CA 93106 and the Speech Communications Research Laboratory, Inc., Santa Barbara, CA 93109.

J. D. Markel is with the Speech Communications Research Laboratory, Inc., Santa Barbara, CA 93109.