Behavior Research Methods, Instruments, & Computers
1993, 25 (1), 67-75

Automatic syllabication for an on-line reading tutor

RICHARD L. VENEZKY and MANDAYAM SURAJ
University of Delaware, Newark, Delaware

To assist limited-proficiency readers, an on-line reading tutor has been developed with a number of word-level aids. One of these, called syllabication, divides printed words into syllables that are then displayed and pronounced one at a time. The algorithm that divides words into syllables takes as input a printed word and its DECtalk-generated pronunciation. Divisions are made first in the pronunciation, using a consonant cluster tree to decide when more than one possible break can be made between two vowels. Rules that consider particular prefixes, suffixes, and stress and vowel patterns then select the most likely division. The resulting system performs above 99% correct for running text, but it fails on certain types of compound words, which can be handled correctly only through exception lists.

Computer-mediated text has been effective in improving the comprehension scores of both disabled and normal readers (Olson, Foltz, & Wise, 1986; Reinking, 1988). However, these results appear to depend on the availability of on-line reading assistance accompanying the computer-mediated text. For example, Reinking found no comprehension differences between good and poor fifth and sixth graders for printed as opposed to computer-mediated text when the latter had no feedback aids. For computer-mediated text with feedback options, however, comprehension scores were significantly higher. A second critical factor appears to be that the computer mediation system, to be effective, must influence cognitive processing of the text (Daniel & Reinking, 1987; Duchastel, 1986). In Reinking (1988), students spent significantly more time reading short expository passages in the feedback-options condition than in either the no-feedback-aids condition or the print condition. In contrast with these results, Kunz, Schott, and Hoverkamp (cited in Reinking, 1988, p. 495ff.) found no comprehension differences for college students between printed texts and computer-mediated texts with assistance. However, no reading-time differences were found either, indicating that the on-line subjects may not have shifted to a more reflective "study mode."

In the majority of the computer-mediated text studies cited, feedback aids have involved additional text displays that give definitions of difficult words, easier versions of the text or of one of its sections, or comprehension hints. Speech feedback has been used in a small number of studies (e.g., Wise et al., 1989), generally with promising results. Wise et al. found that segmented feedback of displayed words, when accompanied by the spoken counterparts, improved the phonological decoding skills of disabled elementary level readers.

In the teaching of children and adults to read, decoding or phonics serves an important role regardless of the method used for the initial introduction of the task. In whole-word methods, generally a small (20-50) sight-word vocabulary is taught first, in simple, meaningful passages. Then decoding patterns are introduced. Even in the "whole-language" methods, decoding is eventually introduced. Phonics methods generally begin directly with letter-sound correspondences from which single-syllable words can be built; these words are then used to make simple sentences, as, for example, "The cat sat on the mat" (Adams, 1991). At some point in the sequencing of tasks for each of these reading methods, multisyllabic words are introduced, and strategies are taught for breaking them into their constituent parts. This process, called *syllabification* or *syllabication* (which will be preferred here), is the focus of the work reported in this paper.[1]

An experimental on-line reading system has been developed at the University of Delaware to assist both children and adults in acquiring basic reading skills. Texts are entered into the system by an optical scanner (and character recognition system), or by keyboard, or directly from a diskette or remote source. A student can select a text for reading and then be prompted to read the displayed material word by word, phrase by phrase, or sentence by sentence. The system can also read any of these units aloud to the student, using the DECtalk system to translate from standard orthography to speech (*DECtalk*, 1985). For the phrase-by-phrase reading mode, a rule-based system developed at the University of Delaware divides texts into reading phrases. The student can also ask for the definition of a word, which is supplied from an included dictionary, for the pronunciation of a word, or for a syllable-by-syllable decomposition of a word with accompanying pronunciation. The system described here is called whenever a student asks for such syllable-by-syllable assistance, here called *decoding assistance*. When decoding assistance is requested, the system opens a window and displays the word in large type. Then, the word is pronounced, followed by the pronunciations of the constituents of the word. As the constituents are pronounced, they are displayed left to right with a single-space underline separating them (e.g., de_cod_ing). The problem that is discussed here is how to segment a polysyllabic word, given only its spelling and pronunciation (with the latter obtained from the former via DECtalk's internal rule system).

The work reported here was supported by a grant from Xerox Imaging Systems, a subsidiary of the Xerox Corporation. Both R.L.V. and M.S. are in the Department of Computer and Information Sciences at Delaware. Correspondence should be addressed to R. L. Venezky, Department of Educational Studies, Willard Hall Education Building, Room 201, University of Delaware, Newark, DE 19716-2920 (e-mail: venezky@dewey.udel.edu).

Syllabication, or the segmentation of a word into syllables, is not the same process as hyphenation. The word *difference*, for example, is divided dif_fer_ence for end-of-line hyphenation but diff_er_ence for reading aloud (i.e., syllabication). Hyphenation rules take into account visual image as well as spoken syllables. Therefore, single-letter syllables are not allowed at the beginning or end of a word, and doubled consonants, even though in most cases standing for a single sound, are divided across syllables. Single consonants between vowels are often placed with the following vowel, even though they might be pronounced with the first. Thus, the *n* in petunia is placed with the final syllable for end-of-line hyphenation (pe_tu_nia) but with the next to last syllable for syllabication /pə'tun jə /.[2] In general, American lexicographers give less emphasis to etymology than do British lexicographers in determining end-of-line divisions. For complete rules from the British standpoint, see *Hart's Rules for Compositors and Readers* (Oxford, 1983); for the American position, see the *Style Manual* of the U.S. Government Printing Office (1973).

APPROACHES TO SYLLABICATION

One approach to syllabication is through dictionary lookup. A dictionary or at least the spellings and pronunciations from such a dictionary might be kept on line, along with an appropriate lookup scheme (e.g., B-tree, letter tables).[3] American dictionaries, in contrast to most British dictionaries, tend to include syllable divisions within the pronunciations. The dictionary approach has a number of drawbacks, however. First, dictionaries are relatively expensive to license and could, therefore, force the price of a literacy workstation above that which most adult literacy programs can afford. A second problem is that even with clever compression schemes, the storage of 100,000+ words (spelling and pronunciation for each) requires either a hard disk or a high-capacity diskette drive.

A third problem is that dictionaries give pronunciations of basic forms but generally not of all common affixed forms. Therefore, even with an on-line dictionary, algorithms would be needed to determine the syllable divisions for such words as *running*, *boyishly*, and *churches*, none of which appears in the most common adult dictionaries such as *Webster's Ninth New Collegiate* (Mish, 1983), *American Heritage* (Morris, 1982), or *Random House* (Flexner, 1987). The rules for syllabicating these words are not overly complex, given the divisions for the base forms from a dictionary plus the DECtalk pronunciations for the complete forms, but this will not solve the problem of dividing words that do not occur at all in common dictionaries. Whether these excluded forms are recent technical terms, place-names, personal names, trade names, or whatever, they could be crucial to text understanding.

A further problem is that common dictionaries do not always agree on where syllable boundaries should be placed, as Table 1 demonstrates. Also, syllable breaks in dictionaries sometimes do not correspond to morphological breaks. For example, *boiler* consists of two morphemes or meaningful parts: boil + er. For pedagogical purposes, we want to display this word as boil_er; however, all common dictionaries divide it as boi_ler. Similarly, *walking* is divided as wal_king rather than as walk_ing, which makes the meaningful components of the word more obvious.

For all of these reasons, a project was initiated to develop algorithms for syllabication. The results of this work, which we feel are acceptable for a pedagogical system, are presented below.

SYLLABLES AND SYLLABLE DIVISIONS IN SPEECH

The syllable in American English speech is an elusive entity, sitting somewhere between the chimera and the California condor on the rareness scale. Linguists agree that syllables exist in speech, but they do not have a testable procedure for defining them. Most can agree on the number of syllables within a given word and therefore on the syllable peaks, which are usually vowels, but even here some disagreement exists, particularly for (1) syllabic consonants (e.g., *prism*), (2) /l/ when preceded by a high front vowel as in *feel*, (3) /r/ when following certain vowels (e.g., *fire*, *hour*), and (4) certain vowel sequences (e.g., *idea*). Some linguists count one syllable in *prism*, *feel*, and *fire*, while others, using exactly the same pronunciations, assign two syllables to each. Nevertheless, these disagreements are limited to a

Table 1
Syllable Divisions Extrapolated from Pronunciations in the *American Heritage Dictionary* (AH), *Webster's Ninth New Collegiate Dictionary* (NC9), and the *Random House Dictionary of the English Language* (2nd ed., unabridged) (RH)

	Dictionary		
Word	AH	NC9	RH
bronchitis	bron_chi_tis	bron_chit_is	bron_chi_tis
cider	ci_der	cid_er	ci_der
irradiate	i_rra_di_ate	irr_ad_i_ate	i_rra_di_ate
orangeade	or_an_geade	or_an_geade	or_ange_ade
orphanage	or_pha_nage	orph_a_nage	or_pha_nage
talkative	ta(l)_ka_tive	ta(l)_kat_ive	ta(l)_ka_tive

well-defined set of problems. But counting syllables in a word is not the same as drawing sharp divisions between such entities.

Studies of language processing have generally assigned a role to the syllable in lexical access (e.g., Taft, 1979; Taft & Forster, 1976) and in word games (Treiman, 1983). Treiman, for example, found evidence for analysis of spoken syllables into an *onset* and a *rime*, but not for further analysis of the rime component into a *peak* and a *coda*. (In an earlier study, MacKay, 1972, argued for a hierarchical internal structure for the syllable, based on synonymic intrusions made by German speakers.)

What makes syllabication in a visual display so difficult is that in normal speech, divisions between syllables (and between words) usually do not exist. A word like *manner* is normally articulated with no clear break in vocal energy between the peaks of the two syllables. Somewhere in the articulation of the /n/, a dip in vocal energy occurs, but this is a diminution in output, not a complete stoppage. Kurath (1964) claims that syllable boundaries occur at points of least prominence (sonority), a position that Ladefoged (1975) among others disputes. Ladefoged asserts that neither the speech stream nor the speech-producing system can serve as an unambiguous basis for syllable divisions.

Where consonant clusters that can neither end nor begin a (spoken) syllable occur between vowels, a syllable boundary must occur (e.g., *admire*), as it must between a sequence of two vowels (e.g., *media*). Certain types of vowels do not ordinarily end syllables in English speech; therefore a single consonant that occurs between such a vowel and another vowel usually belongs to the first vowel to "close" the syllable (e.g., *inner*, *ladder*). In a small number of cases vowel quality and syllable boundaries covary. For example, the /r/ in words like *hero* and *zero* can be pronounced at the end of the first syllable or the beginning of the second. If it occurs at the end of the first syllable, the vowel sound in that syllable will tend to be /ɪ/ as in *bit*. If /r/ starts the second syllable, the vowel in the first syllable will be /i/ as in *beet*.

Finally, as pointed out above, the places where syllable boundaries generally occur in normal speech may not be where boundaries should be shown if teaching a word's pronunciation is the goal. To assist in recognizing and pronouncing a word like *building*, for example, the two meaningful constituents should be shown even though most speakers would make a syllable break between the /l/ and the /d/. Since syllable boundaries by themselves are not indicators of word meanings, they often can be arbitrarily placed. *Firing* can be spoken as fi_ring or as fir_ing (or with no clear syllable break) without sounding peculiar and without affecting the identity of the word for the listener. Although English speakers tend to ignore word derivation in making syllable breaks in speech, exceptions are made for certain prefixes. For example, *distract* and *destroy* begin with identical pronunciations and both have second syllable stress. *Destroy* follows the division pattern found for almost all words in which the intervocalic cluster /str/ follows an unstressed vowel; that is, /str/ remains intact as the onset of the vowel that follows. In *distract*, however, the prefix *dis* /dis/ remains intact, yielding dis_tract. Similarly, the general rule that would, under selected conditions, assign an intervocalic consonant to the vowel that follows, is suspended for the prefix *sub* whenever the root that follows exists as a separate word (e.g., sub_ar_id). In *subaltern*, the root is not recognized as a separate word (even though it is an archaic form of the noun *alternate*); therefore, the word divides as su_bal_tern, with stress on the second syllable. (But note that *subalternate* divides after the /b/ as an adverb but before the /b/ as an adjective.) Other prefixes that tend to remain intact in similar contexts are *con-*, *im-*, *in-*, and *un-*.

Trans- also follows this pattern, but ambiguities arise when the root that follows begins with /s/, as in *transcribe* and *transpire*. In connected speech, a single /s/ is articulated, connecting the two syllables without a break. The quandary this presents for lexicographers is reflected in the different treatments given *trans-* words in modern dictionaries. *Webster's Ninth New Collegiate* (Mish, 1983) keeps *trans-* intact without exception, leading to such implausible pronunciations as trans_pire, trans_cribe, and trans_is_tor (the last being a compound of *trans[fer]* and *[re]sistor*, made up in the late 1940s when transistors were first developed). *The Random House Dictionary* (Flexner, 1987) prefers the more natural divisions tran_spire, tran_scribe, and tran_sis_tor. Although neither is exactly accurate, the break after the /n/ is more natural in these cases. Our solution is to keep *trans-* intact except where the /s/ begins a recognizable component. In these cases, a break is made before /s/ rather than after. To add to the complexity of morphological recognition, English appears to have a melting-pot principle for pronunciations of multisyllabic words, but this principle is not applied evenly across the lexicon. *Nothing*, a compound of *no+thing*, has been naturalized and divides now as noth_ing as opposed to the earlier no_thing. In contrast, *something* has resisted naturalization and remains as some_thing. *Kilometer* represents a word in transition, breaking either as kil_o_me_ter or as ki_lom_e_ter.

MISPRONUNCIATIONS

DECtalk 3.1, the speech synthesis system used here, is a remarkably accurate text-to-speech system, given the complexities of English grapheme-phoneme correspondences (Bruckert, Minow, & Tetschner, 1983; Klatt, 1980). It also has a high degree of intelligibility for both children and adults, not differing significantly from recorded human speech in some tests (e.g., Olson et al., 1986; but cf. Greene & Pisoni, 1988). Nevertheless, decoding errors do occur, some of which lead to syllabication errors. Among these are the following:

1. Since English allows almost any form of word compounding, the failure to recognize certain words as compounds leads to mispronunciations that affect syllable di-

visions. For example, *woodshed* is not in DECtalk's compound list; therefore it is treated as *woodsh* + verbal past tense marker *ed*, yielding the pronunciation /wʊdst/. With only one vowel in the pronunciation, the syllabication system generates no syllable divisions.

2. In some cases, vowel sounds are deleted, particularly with the sequence *ua* as in *usually* and *annual* and with medial *e* as in *conference* and *interest*. *Interest*, for example, is divided in_t(e)rest according to the DECtalk pronunciation.[4]

3. In many words in which /r/ or /l/ occurs between a consonant and a vowel (e.g., *titrate*, *cyclic*), an extra syllable is generated. For example, *titrate* is pronounced /ti tə ret/ and is therefore divided as ti_tr_ate.

4. In several words (e.g., *misspell*, *cannot*), the second in a sequence of identical consonants is omitted.

5. An occasional multisyllabic word is not assigned a stress mark (e.g., *having*), and a few are assigned two secondary but no primary stress marks (e.g., *unfailing*). Stress placement errors also occur (e.g., *orogeny*).

6. Since DECtalk uses [r] for the consonant typically spelled *r* and the sequence [rr] for both the stressed and unstressed *r*-colored vowels as in *bird* and *water*, potentially ambiguous sequences of three and four /r/s can occur, as in *mirror* and *manufacturer*. In general, these can be disambiguated, except for unusual compounds like *mirrorrod* (=*mirror*+*rod*).

VISUALIZATION

Solving the phonological issues, however arbitrarily, still leaves for solution a number of problems related to the visual representation of syllable breaks. The letter *x* in any word position other than initial represents two different sounds, as in *exist*, *luxury*, and *taxes*. In all cases when *x* is neither the first nor the last letter in a word, the two sounds it represents are in separate syllables. How then should these words be divided? The decision we have made, although arbitrary, is to include the letter *x* with the syllable in which its first sound (/k/ or /g/) occurs. Thus, *explained* is shown as ex_plained.

A second visualization issue occurs in words in which certain consonants have become, under the influence of a following /ɪ/ or /j/, new sounds that have absorbed the following / ɪ / or /j/. This process, called *palatalization*, is represented by the /š/ in *official*, the /ž/ in *treasure*, the /č/ in *bastion*, and the / ǰ / in *educate*. Of the examples just given, only *official* presents a visualization problem. Our solution is to include the palatalizing letter, if possible, with the consonant that it affects. This gives o_ffici_al but ed_u_cate. Related to this problem is the general issue of silent letters. An initial or final silent letter can be attached to its adjacent syllable without stirring up undue attention, but what should be done with silent vowels within a word, as in *business*, *arsenal*, and *arsenic* (assuming two-syllable pronunciations for the latter two words)? In all cases where silent letters can be detected, except for final, silent *e*, we enclose the letters in parentheses: (k)nife, ri(gh)ts, of_(t)en.

SYLLABICATION ALGORITHM

The algorithm described here uses PROLOG rules to parse a word into an acceptable sequence of syllables, as defined by the criteria given above. The flowchart for the algorithm is shown in Figure 1.

The syllabication algorithm operates on pronunciations rather than on spellings. This is done because the phonological properties of vowels (checked, free, schwa, stressed, unstressed, etc.) and some aspects of consonant sounds that are essential for determining syllable divisions are not apparent from their spellings alone. The syllabication process is therefore based largely on the phonemic equivalent of the word (i.e., a coded form of the pronunciation of the word). This phonemic equivalent of the word is obtained through DECtalk, which can be set to a mode called "log phoneme." In this mode, the system returns the phonemic equivalent of an orthographic string. This phonemic version of the word is in "two-letter" codes, which are then converted to single-letter codes for flexibility and ease of processing.

For example, if the word *mechanical* is sent to DECtalk, the phonemic string [mixk'aenixkel] is returned.[5] This in turn is converted by the system discussed here to a list of coded characters "['m','X','k','A','n','X','k','F','l']." The codes used by DECtalk and the corresponding ones to which they are converted are shown in Table 2.

DECtalk also returns stress markers (primary stress, ', and secondary stress, `) before stressed vowels, and returns an asterisk ('*') between the two parts of a compound word (like *hothead*), whenever it identifies one. The stress markers are stripped off, but their locations are retained for use by certain rules. The '*' is used for syllabication and will be explained later. Syllabication is done on the list of code characters thus obtained.

For example, "['m','X','k','A','n','X','k','F','l']" is divided into "['m','X'],['k','A','n'], ['X'],['k','F','l']". This is used subsequently to split the word itself into its syllables, me_chan_i_cal. The terms *consonant* and *vowel* will henceforth refer to phonemes. For example, the word *silhouette* with phonemic representation in DECtalk codes [sihloweht] has three consonants, namely, [s], [l], and [t], and three vowels, namely [ih], [ow], and [eh]. Given the phonemic equivalent of a word, the number of syllables in the word equals the number of vowels in the phonemic version. Each vowel forms the nucleus of a syllable and can have (optionally) consonants on either of its sides. Between every pair of contiguous vowels, a syllable boundary must occur. However, whenever one or more consonants occur between the two vowels, the syllable boundary might occur at any of the possible break points. Thus, as explained above, the task in automatic syllabication is to determine boundary placement whenever one or more consonants occur between two vowels.

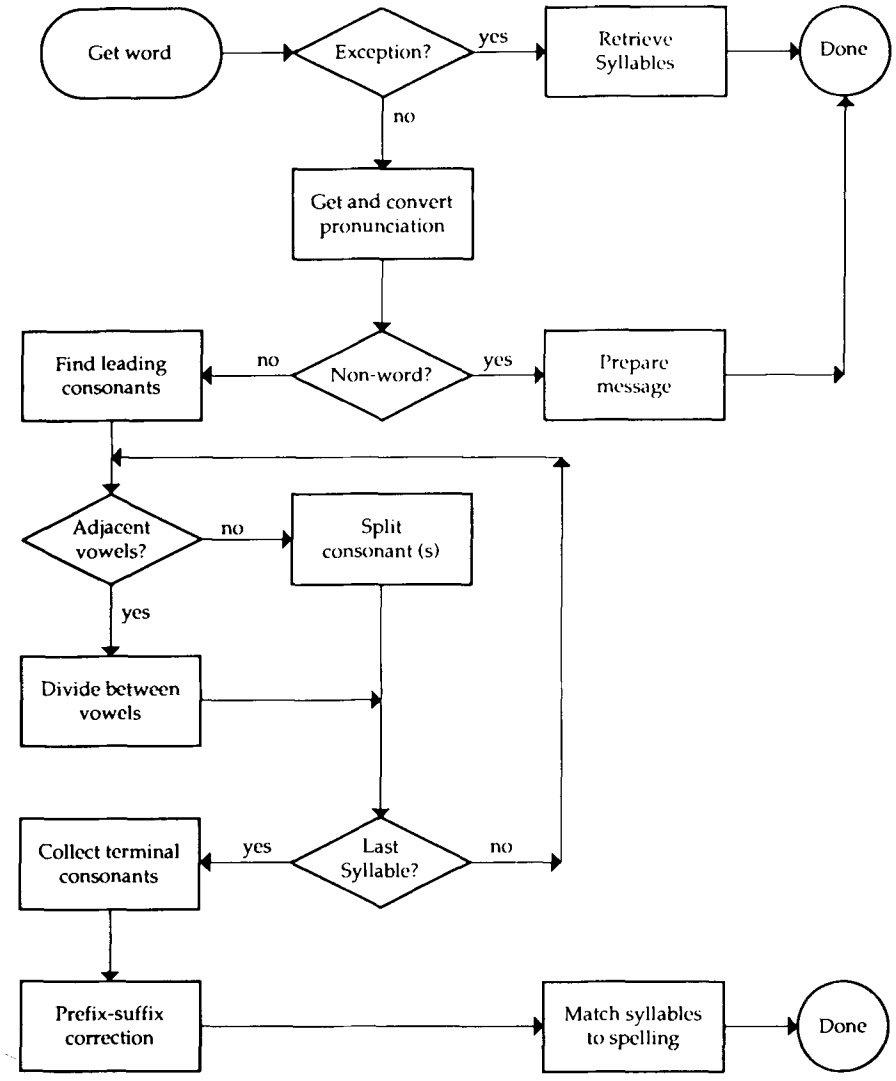

Figure 1. Flowchart for syllabication algorithm.

RULES

Step 1.1

The word is checked against an exception list. If the word is matched, the syllable breaks are extracted from the list. This list is used mainly to detect certain compound words like *everything*, but it also includes words which DECtalk mispronounces, like *misspelling*.

Step 1.2

A check is made to detect abbreviations and stand-alone letters (e.g., DECtalk reads the letter *w* as *double u*). Abbreviations with one, two, and three letters that have corresponding phonemic representations that are at least three units longer than their spellings are assumed to be abbreviations or spelled letters.

Step 2

All leading consonants and the first vowel are collected as part of the first syllable. For the word *streamer* (['s','t','r','I','m','R']), this gives "[['s','t','r' . . .] . . .]".

Step 3

Find all sequences of two consecutive vowels and split between each pair. For example, in the word *skiing*, coded as "['s','k','I','X','N']," split between [I] and [X]. For sequences of VC . . . V, where V represents a vowel and C . . . represents one or more consonants, the consonant parsing trees shown in Table 3 are checked for potential break points. Then, if more than one potential break point exists between two vowels, the rules that follow are checked and the first one that applies is used. Each case is based on the number of consonants that exist between

Table 2
DECtalk Pronunciation Symbols

English Example	DECtalk Alphabet	1-letter Code	IPA Code	English Example	DECtalk Alphabet	1-letter Code	IPA Code
father	aa	a	a	bet	eh	e	ɛ
bit	ih	i	ɪ	bought	ao	o	ɔ
cute	yu	u	ju	bat	ae	A	æ
bake	ey	E	e	beat	iy	I	i
boat	ow	O	o	lute	uw	U	u
bar	ar	1	ar	bear	er	2	er
beer	ir	3	ɪr	bore	or	4	or
poor	ur	5	ur	bite	ay	y	aɨ
boy	oy	Y	ɪɪ	book	uh	H	ʊ
bout	aw	W	au	about	ax	x	ə
kisses	ix	X	ɪ	bird	rr	R	ʌr
but	ah	B	ʌ	singer	rr	R	ʌr
bet	b	b	b	debt	d	d	d
fin	f	f	f	gift	g	g	g
head	hx	h	h	gin	jh	J	j
cat	k	k	k	let	l	l	l
met	m	m	m	net	n	n	n
sing	nx	N	ŋ	pin	p	p	p
red	r	r	r	sit	s	s	s
shin	sh	S	š	chin	ch	C	č
test	t	t	t	thin	th	T	θ
this	dh	D	ð	vest	v	v	v
wet	w	w	w	yet	y	j	J
zoo	z	z	z	leisure	zh	Z	ž
usually	e	F	ə				

a pair of vowels. The trees in Table 3 are represented in a compact parenthesized form. The entry for /f/ (line 4) is read as follows:

/fj/, /fl/, and /fr/ can occur in syllable-initial position (position 1). The cluster /fts/ can occur in syllable final position (position 2), as can the cluster /ft/ (# stands for syllable boundary). And /f/ by itself can occur in either syllable-initial or syllable-final position.

Table 3
Consonant Parsing Trees

1. b [j(1), l(1), r(1), z(2), #(1,2)]
2. ch [l(1), r(1), #(1,2)]
3. d [j(1,2), r(1), w(1), z(1), #(1,2)]
4. f [j(1), l(1), r(1), t [s(2), #(2)], #(1,2)]
5. g [l(1), r(1), z(2), #(1,2)]
6. h [j(1), w(1), #(1)]
7. j [#(1)]
8. j [#(1,2)]
9. k [j(1), l(1), r(1), s [t(2), th(2), #(2)], t [s(2), #(2)], w(1), #(1,2)]
10. l [b(2), d [z(2), #(2)], ch(2), f [th(2), #(2)], k(2), m(2), p [s(2), #(2)], sh(2), t [s(2), #(2)], th(2), z(2), #(1,2)]
11. m [j(1), f(2), p [s(2), t [s(2), #(2)], #(2)], t(2), th(2), z(2), #(1,2)]
12. n [ch(2), d [th(2), z(2), #(2)], j(1), s [t(2), #(2)], t [s(2), #(2)], th [s(2), #(2)], z(2), #(1,2)]
13. ng [k [s(2), #(2)], s[t(2), #(2)], th(2), z(2), #(2)]
14. p [j(1), l(1), r(1), s(2), t(2), th(2), #(1,2)]
15. r [s(2), z(2), #(1,2)]
16. s [j(1), k [s(2), r(1), w(1), #(1,2)], l(1), m(1), n(1), p [l(1), r(1), #(1,2)], r(1), t [r(1), s(2), #(1,2)], w(1), #(1,2)] s[e(1), m(1), n(1), r91), t(1,2),#(1,2)]
17. t [j(1), r(1), s(2), #(1,2)]
18. th [r(1), s(2), #(1,2)]
19. dh [m(2), z(2), #(1,2)]
20. v [j(1), z(2), #(1,2)]
21. w [#(1)]
22. z [m(2), #(1,2)]
23. zh [#(1,2)]

Note—# = syllable juncture; 1 = syllable initial; 2 = syllable final.

Case 1: One consonant. 1.1. If the first vowel is stressed and the second is a schwa /ə/, then group the consonant with the stressed vowel. For example, for the word *totemism* (''['t','O','t','x','m','x','z','x','m']''), ['O','t','x'] splits as [. . .[..'O','t'],['x'..]. . .]. To determine whether the vowel [O] is stressed or not, the copy of the original phonemic version of the word (with stress markers) is searched for a stress mark before [O].

1.2. If the first vowel is an [E] or [I] and the consonant is [r], then split as [. . .[..'E','r'],[. . .]. . .].

1.3a. If the first vowel is stressed and the consonant is [r], split as [. . .[..V1,'r'],[V2..]. . .].

1.3b. If the first vowel is unstressed and the consonant is [r], split as [. . .[..V1],['r',V2..]. . .].

1.4a. If the first vowel is short[6] and stressed, then group the consonant with it.

1.4b. If the first vowel is short and unstressed, then group the consonant with the second vowel.

1.5. If the first vowel is long[7] or a schwa, group the consonant with the second vowel.

Case 2: Two consonants. 2.1. If the first vowel is a schwa, group the two consonants with the second vowel.

2.2. If possible, choose according to the routine for splitting consonant clusters. This routine chooses splits for consonant clusters in the following order:

<1> Split before cluster if it is one of [br fr pr kr gr dr tr bl fl gl kl pl].

<2> Split between the two consonants if the first can possibly end a syllable and the second can possibly start a syllable.

This can be determined from the tree expressions in Table 3. For example, in the word *artist* (['a','r','t','X','s', 't']), the consonant cluster ['r','t'] between the vowels [a] and [X] can split between [r] and [t]. This is because [r] can terminate a syllable and [t] can begin one. (Note, however, that [rt] cannot begin a syllable).

<3> Split after the two consonants similarly.

<4> Split before the two consonants.

Case 3: Three consonants. 3.1. If the first vowel is a schwa, group the three consonants with the second vowel if they can start a syllable.

3.2. If the first vowel is neither short nor a schwa and is not stressed and the consonants are ['s','t','r'], then split before the consonants.

3.3. Split the consonant cluster according to the following order:

<1> Between the first and second consonants.

<2> Between the second and third consonants.

<3> Between the first vowel and the first consonant.

Case 4: Four consonants. 4.1. Split the consonant cluster according to the following order:

<1> Between the first and second consonants.

<2> Between the second and third consonants.

<3> Between the third and fourth consonants.

<4> Between the fourth consonant and the second vowel.

If the consonant cluster between two vowels is not handled by the rules above, use a default rule to split it. Such clusters can result from compound words. The default rule used is to search backwards from the last consonant in the cluster and find the longest possible sequence of consonants before which a break can occur according to the tree expressions of Table 3. Then, break the consonant cluster before that sequence of consonants.

Step 4:

Collect all terminal consonants as part of the last syllable.

Compound Words

DECtalk returns a morpheme boundary marker in the phonemic representation of some words like *throughout* [thruw*'awt] and *talk* [*taok]. If the '*' occurs at the beginning of the phoneme string, it is discarded. When it occurs within the phoneme string, it is used as an indicator for a syllable break at its position. This is done by treating '*' as a vowel. Doing this ensures that it occurs in some syllable along with zero or more consonants. As soon as this syllable is formed (during the execution of the algorithm above), all consonants preceding the '*' are chained to the previous syllable and all consonants succeeding the '*' are grouped to form part of the next syllable. The '*' itself is then deleted.

SECOND STAGE PROCESSING

The processing just described gives a preliminary parsing of the word. This is, however, not always the correct parsing because of the influence of prefixes and suffixes. The second stage handles these problems. In the second major stage, the spelling of the word and its parsed pronunciation are matched. This is not a simple task because of the nature of English orthography. Silent letters (as in *hour* [awr], *business* [b'ihznixs]), extra sounds (as in *excuse* [ixksk'yuz]), and unusual correspondents (as in *colonel* [k'rrnel]) have to be handled. (See Venezky, 1970, for a fuller description of English orthography.) The steps in this stage are the following:

Step 1

Separate suffixes and prefixes. The suffixes that are separated are *-ing*, *-asm*, and *-ism*. Other suffixes are parsed automatically as a result of the rules used. The prefixes that are separated are *sub-*, *trans-*, *self-*, and *dis-*. The *sub-*, *trans-*, and *self-* prefixes are handled in a similar way. For example, for *trans-*, the spelling of a word is first tested for initial *trans-*. Then, the phonemic equivalent is tested for initial ['t','r','a','n','s'] or ['t','r','a', 'n','z']. (Note that *s* in *trans-* can sound like [s] or [z] as in the words *transcribe* ([traenskr'ayb]) and *transect* ([traenz'ehkt].) If these test true, an exception list is searched for the word. If the word is found in the list, *trans-* is not a prefix in that word and parsing is done according to the other rules; otherwise *trans-* is separated before parsing continues.

Dis- is separated differently. In this case, the initial part of the phonemic representation of the word is first checked for ['d','X','s'] (i.e., [dixs]). Then the initial part of the spelling is checked for *dis-*. If both of these are true and the word does not belong to an exception list, the *dis-* prefix is separated; otherwise, parsing is done according to the other rules.

Step 2

Try to match the next string of spelling consonants with a single phoneme consonant, starting with the longest possible string and continuing with decreasing order of lengths (with three, then two, and lastly one consonant), until a match occurs. The process of matching is based on a set of PROLOG *facts* that list all possible spellings in English that can result in a particular phoneme. For example, the phoneme /f/ can be spelled *gh*, *ph*, *ff*, or *f* as in *cough*, *phase*, *coffee*, and *free*. Therefore, four PROLOG facts are required:

```
sounds_like("gh", 'f').
sounds_like("ph", 'f').
sounds_like("ff", 'f').
sounds_like("f", 'f').
```

Thus, sound-spelling correspondences are stored for all consonants. For unique correspondences such as the /r/ in *colonel*, the vowel environment is included with the spelling (*olo*); since vowel correspondences are not compared, the correspondence of /r/ to *olo* is detected in *colonel* but not in *hologram*.

Vowels are matched sequentially, sound to spelling, without regard for actual correspondences. Any single vowel sound is matched to any sequence of one or more

vowel letters. If two vowel phonemes occur together and two vowel spellings are available before the next consonant spelling or juncture, a one-to-one correspondence is made. If two consecutive vowel phonemes have to be matched to three consecutive vowel spellings, the following spelling list is searched for the first match to either the first and second letters or the second and third letters: *ou*, *au*, *oi*, *ai*, *ae*, *ee*, *ei*, *ow*. If a match is found, the spelling string is parsed accordingly; otherwise, a split is made after the second vowel letter.

When a type mismatch occurs between the phonemic string and the spelling string—that is, when a vowel spelling is found where a consonant is sought, or vice versa—the first two unmatched letters of the spelling string are compared with potential silent-letter strings (e.g., *gh*, *ue*). If matched, they are stripped off and the matching process continues. If this fails to resolve the mismatch, the next available single letter is tested similarly. If this fails, all letters whose type (vowel, consonant) differs from that of the phoneme in question are marked as silent.

Step 2 is repeated until the entire list of phonemes is matched. The completely parsed spelling and phoneme strings are then passed to a display procedure, where the syllables are displayed and pronounced one by one.

CONCLUSIONS

For running text, the algorithm presented here performs above a 99% accuracy level and is adequate, therefore, for a practical reading system. It will continue to fail, nevertheless, on certain compound words such as *woodshed*, which DECtalk grossly mispronounces, and on words such as *children* and *prism*, where DECtalk inserts an extra vowel (/čɪldərən/ and /pərɪzɪm/). The latter cases could be detected and edited; future work on the syllabication system will be focused on this issue. A second problem stems from the desire to isolate suffixes like *-ing* and *-er*. For *-ing*, a simple segmentation rule is adequate. But *-er* was found to be a less tractable problem. Some *-er* words like *rancher* and *shuffler* do not divide naturally with *-er* as a separate syllable. In many others, morphemic *-er* cannot be clearly separated from nonmorphemic *-er* (e.g., *sleeker* vs. *panzer* and *panther*). Consequently, no special attempt is made to isolate morphemic *-er*.

The program is available from the author at no cost. Send a 5.25-in. diskette, a return address label, and a diskette mailer or, for an e-mail copy, send a request to the author at venezky@dewey.udel.edu.

REFERENCES

Adams, M. J. (1991). *Beginning to read*. Cambridge, MA: MIT Press.

Bruckert, E., Minow, W., & Tetschner, W. (1983). Three-tiered software and VLSI aided developmental system to read text aloud. *Electronics*, **56** (8), 133-138.

Daniel, D. B., & Reinking, D. (1987). The construct of legibility in electronic reading environments. In D. Reinking (Ed.), *Reading and computers: Issues for theory and practice* (pp. 24-39). New York: Columbia University, Teachers College Press.

DECtalk DTC03 text-to-speech system owner's manual (1985). (Rep. EK-DTC03-OM-001). Maynard, MA: Digital Equipment Services.

Duchastel, P. (1986). Computer text access. *Computer Education*, **10**, 403-409.

Flexner, S. B. (Ed.) (1987). *The Random House dictionary of the English language* (2nd ed.). New York: Random House.

Greene, B. G., & Pisoni, D. B. (1988). Perception of synthetic speech by adults and children: Research on processing voice output from text-to-speech systems. In L. E. Bernstein (Ed.), *The vocally impaired: Clinical practice and research* (pp. 206-248). Philadelphia: Grune & Stratton.

Klatt, D. H. (1980). Software for a cascade/paralleled format synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

Kurath, H. (1964). *A phonology and prosody of modern English*. Ann Arbor, MI: University of Michigan Press.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, **3**, 210-227.

Mish, F. C. (Ed.) (1983). *Webster's ninth new collegiate dictionary*. Springfield, MA: Merriam-Webster.

Morris, W. (Ed.) (1982). *The American Heritage dictionary of the English language* (2nd college ed.). Boston: Houghton Mifflin.

Olson, R., Foltz, G., & Wise, B. (1986). Reading instruction and remediation with the aid of computer speech. *Behavior Research Methods, Instruments, & Computers*, **18**, 93-99.

Oxford University Press (1983). *Hart's rules for compositors and readers at the University Press, Oxford* (39th ed.) Oxford: Author.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly*, **23**, 484-498.

Taft, M. (1979). Lexical access via an orthographic code: The basic orthographic syllable structure (BOSS). *Journal of Verbal Learning & Verbal Behavior*, **18**, 21-39.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning & Verbal Behavior*, **15**, 607-620.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition*, **15**, 49-74.

United States Government Printing Office (1973). *Style manual* (rev. ed.). Washington, DC: Author.

Venezky, R. L. (1970). *The structure of English orthography*. The Hague: Mouton.

Wise, B., Olson, R., Anstett, M., Andrews, L., Terjak, M., Schneider, V., Kostuch, J., & Kriho, L. (1989). Implementing a long-term computerized remedial reading program with synthetic speech feedback: Hardware, software, and real-world issues. *Behavior Research Methods, Instruments, & Computers*, **21**, 173-180.

NOTES

1. In Wise et al. (1989) and Olson et al. (1986), this process is referred to as syllable segmentation.
2. IPA symbols are used here for phonological representation, enclosed within slanted lines. Table 2 gives a list of IPA symbols with equivalent DECtalk codes.
3. Olson et al. (1986) hand coded syllable boundaries for their stimulus materials. Although this is adequate for experimental purposes, it would not be practical for a general-purpose instructional system.
4. Silent letters, except for final, silent *e*, are enclosed in parens.
5. DECtalk phonological codes are enclosed in square brackets.
6. The short vowels are /æ, ʊ, ɛ, ʌ, ɪ/.
7. The long vowels are all vowels that are neither short nor schwa (/ə/).

APPENDIX
Sample Output From the Syllabication Algorithm

DecTalk	Spelling	Syllables
axbl'eyz	ablaze	a_blaze
axkl'aymaxtayz	acclimatize	a_cclim_a_tize
'axdmihnixstr'eyshixn	administration	ad_mi_ni_stra_tion
b'oyixshliy	boyishly	boy_ish_ly
braanxk'aytixs	bronchitis	bron_chit_is
b'ihldixnx	building	build_ing
b'ihznixs	business	bus(i)_ness
s'aydrr	cider	ci_der
sixvihlixz'eyshixn	civilization	ci_vi_li_za_tion
dehklaxr'eyshixn	declaration	de_cla_ra_tion
dixstr'oy	destroy	de_stroy
d'ehvaxsteyt	devastate	dev_a_state
dixskixnt'ihnyu	discontinue	dis_con_tin_ue
'ehksixkyut	execute	ex_e_cute
ixgz'ihst	exist	ex_ist
f'ehrnaxs	fairness	fair_ness
hx'iyrow	hero	her_o
hx'aanrreyriy	honorary	hon_or_ar_y
hx'aydraxjhixn	hydrogen	hy_dro_gen
ayd'iyax	idea	i_de_a
'ihmixteyt	imitate	im_i_tate
'ihmaxnixnt	imminent	imm_i_nent
ixmp'owrtixnt	important	im_por_tant
lihlixp'yushixn	lilliputian	li_lli_pu_tian
l'aajhixk	logic	log_ic
l'uwkrr	lucre	lu_cre
l'ahkshrriy	luxury	lux_ur_y
mixk'aenixkel	mechanical	me_chan_i_cal
m'ehtrixk	metric	me_tric
owr'ihjhixnel	original	o_rig_i_nal
praxr'ehkwixsixt	prerequisite	pre_re_qui_site
rixs'iyt	receipt	re_cei(p)t
s'ehlflixsliy	selflessly	self_less_ly
sahbkyut'eyniyixs	subcutaneous	sub_cu_ta_ne_ous
traenzaxbstaenshiy'e	transubstantiation	tran_sub_stan_ti_a_tion
'ahnf'eylixnx	unfailing	un_fail_ing
v'aybraxnt	vibrant	vi_brant
v'ayrixs	virus	vir_us
z'iyrow	zero	zer_o

(Manuscript received August 6, 1991;
revision accepted for publication May 26, 1992.)