Computer Speech and Language (1987) **2**, 133–142

The predominance of strong initial syllables in the English vocabulary

Anne Cutler*

MRC Applied Psychology Unit, Cambridge, U.K.

and

David M. Carter

Computer Laboratory, University of Cambridge, U.K.

Abstract

Studies of human speech processing have provided evidence for a segmentation strategy in the perception of continuous speech, whereby a word boundary is postulated, and a lexical access procedure initiated, at each metrically strong syllable. The likely success of this strategy was here estimated against the characteristics of the English vocabulary. Two computerized dictionaries were found to list approximately three times as many words beginning with strong syllables (i.e. syllables containing a full vowel) as beginning with weak syllables (i.e. syllables containing a reduced vowel). Consideration of frequency of lexical word occurrence reveals that words beginning with strong syllables occur on average more often than words beginning with weak syllables. Together, these findings motivate an estimate for everyday speech recognition that approximately 85% of lexical words (i.e. excluding function words) will begin with strong syllables. This estimate was tested against a corpus of 190 000 words of spontaneous British English conversion. In this corpus, 90% of lexical words were found to begin with strong syllables. This suggests that a strategy of postulating word boundaries at the onset of strong syllables would have a high success rate in that few actual lexical word onsets would be missed.

1. Introduction

The recognition of continuous speech necessarily involves a process of segmentation. Recognizers cannot hold in memory a representation of each and every whole utterance they might conceivably be confronted with, because the number of such utterances is infinite. Instead, they must store representations of the discrete units of which utterances may be made up. Consequently, in order to achieve recognition, i.e. to locate the stored representations which correspond to the acoustic forms in a speech signal, the recognizer must in some way segment the signal into its component units, or words.

*Postal Address: MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2 2EF, U.K.

0885-2308/87/030133+10 $03.00/0

The process of segmentation would be simple if the boundaries between words were reliably marked. However, speech researchers have so far failed to find reliable correlates of word boundaries in continuous speech signals. Under these circumstances, there are three alternative courses of action open to speech researchers, whether they are attempting to model the human speech recognition process or to construct an automatic speech recognition system. One is to do no prelexical segmentation at all, but to initiate lexical access attempts only after successful recognition of preceding words. The second is to undertake segmentation, and the initiation of lexical access attempts, on an entirely arbitrary basis (such as every *n* milliseconds; or at every new phonetic segment if a phonetic classification is undertaken). This results in a very large number of fruitless lexical access attempts. The third alternative is to restrict segmentation and consequent lexical access attempts according to some principle of likely high efficiency.

One such principle has been proposed for human speech recognition by Cutler & Norris (1988). The principle is based on metrical prosody, i.e. the rhythmic structure of language. In a stress language like English, syllables can be either strong or weak. Strong syllables always contain full vowels; weak syllables contain reduced vowels (these are usually schwa, but may also be a short form of another vowel, as in the first syllable of *invent* or the second syllable of *barrow*). Cutler & Norris (1988) found that listeners tended to segment nonsense sequences at the onset of strong syllables. They showed this in an experiment in which they presented listeners with bisyllabic initially-stressed nonsense sequences such as *mintayf* (in which the second vowel is strong: [mɪntef]) versus *mintef* (in which the second vowel is schwa: [mɪntəf]); the listeners' response time to detect the embedded real word (in this case, *mint*) was measured. Responses were significantly slower when the second syllable was strong; that is, *mint* took longer to detect in *mintayf* than in *mintef*. This, Cutler & Norris (1988) argued, was because *mintayf* was segmented prior to its second syllable, and detection of *mint* therefore required assembly of speech material across a segmentation position. No such difficulty would arise for the detection of *mint* in *mintef* since a weak second syllable would not trigger segmentation. Subsequent experiments established that the difference was not due to some artifact such as differences in length or loudness of the second syllable, or differences in the way the four phonemes of *mint* had been realised in the two contexts. Cutler & Norris (1988) proposed that listeners use strong syllables as the basis for a segmentation strategy in continuous speech perception; the occurrence of a strong syllable triggers segmentation. The motivation for the strategy is the detection of word onsets and hence the facilitation of lexical access. Strong syllables are taken to be likely word onsets, and a lexical access attempt is initiated at each strong syllable.

The present paper examines the characteristics of the English vocabulary with a view to determining the likely success rate of a strategy such as that proposed by Cutler & Norris (1988). Because the strategy is aimed at the efficient initiation of lexical access, the analysis concentrates on the lexical portion of the vocabulary—lexical items, also known as open class or content words. This portion of the vocabulary includes all nouns, verbs, adjectives and most adverbs. The remainder of the vocabulary consists of grammatical items (or closed-class or function words) such as articles, conjunctions and pronouns. Since such items can only be interpreted in relation to the context in which they occur, their recognition does not involve the same process of access to a stored representation which the recognition of lexical words involves. Indeed, there is psycholinguistic evidence which supports this suggestion that grammatical words are recognized quite differently from lexical words (e.g. Bradley, 1978; Friederici & Schoenle, 1980).

Our investigation consisted of three parts: analysis of the metrical structure of the English vocabulary; comparison of metrical structures with respect to frequency of occurrence; and calculation of the metrical characteristics of a corpus of spoken British English.

2. Vocabulary distribution

The main corpus analyzed was the MRC Psycholinguistic Database (Coltheart, 1981). This contains over 98 000 words, based on the Shorter Oxford English Dictionary; 33 313 of these are phonetically transcribed, using the transcriptions of Jones (1963). Table I shows the metrical characteristics of the initial syllables of the transcribed words, divided into four categories: monosyllables (such as *cat* or *screech*), polysyllables with primary stress on the first syllable (such as *analogue* or *structure*), polysyllables with secondary stress on the first syllable (such as *psychological* or *stampede*) and polysyllables with weak initial syllables (in which the vowel in the first syllable is usually schwa, as in *avoid* or *strabismus*, but may also be a reduced form of another vowel, as in *invent* or *excessive*). Any of the first three categories would satisfy the segmentation strategy proposed by Cutler & Norris (1988); it can be seen that these three categories account for 73% of the words in this database. A similar distribution is found in the computer-readable 20 000-word dictionary of American English based on the *Merriam-Webster Pocket Dictionary*, as shown in Table II. (The larger proportion of monosyllables and slightly lower proportion of words with weak initial syllables presumably results from the fact that this dictionary contains a smaller sample of the English vocabulary than the MRC Database; monosyllables tend to be common words, while words with weak initial

TABLE I. Metrical characteristics of word-initial syllables in the MRC Psycholinguistic Database

	Sum	Proportion
Monosyllables	3906	0·117
Polysyllables with initial primary stress	16 842	0·506
Polysyllables with initial secondary stress	3557	0·107
Polysyllables with weak initial syllable	9008	0·270

TABLE II. Metrical characteristics of word-initial syllables in a 20 000 word dictionary of American English

	Sum	Proportion
Monosyllables	3458	0·173
Polysyllables with initial primary stress	9595	0·481
Polysyllables with initial secondary stress	2434	0·122
Polysyllables with weak initial syllable	4473	0·224

TABLE III. Metrical characteristics of word-initial syllables of lexical words in the MRC Psycholinguistic Database

	Sum	Proportion
Monosyllables	3764	0·114
Polysyllables with initial primary stress	16 810	0·509
Polysyllables with initial secondary stress	3546	0·107
Polysyllables with weak initial syllable	8940	0·270

syllables tend to be more specialized or abstruse and hence more likely to be omitted from a smaller dictionary. Other differences between Table I and II may also in part reflect the reported tendency of American English to favour initial stress more than British English does; *research* and *address* for instance, have weak initial syllables in Table I, strong initial syllables in Table II. The ratio of polysyllables with initial primary stress to polysyllables with weak initial syllables rises from 1.87:1 in Table I to 2.15:1 in Table II.)

Tables I and II do not distinguish between lexical and grammatical words. However, since the MRC Database includes form class information, it was possible to analyze lexical words alone. This analysis is presented in Table III.

Tables I, II and III show that the most common lexical type in English is a polysyllable with initial primary stress. Within this category, the most common type is a bisyllable with a weak second syllable (Carlson, Elenius, Granstrom & Hunnicutt, 1985). This does not imply, however, that the majority of words encountered by a recognizer will be of this type. The investigation of Carlson *et al.*, (1985) which was based on a sample less than a third the size of the MRC Database phonetically transcribed sample, found that monosyllabic types were more common than any polysyllabic type. Since the Carlson *et al.* (1985) sample was chosen on the basis of high frequency of occurrence, it is likely that consideration of frequency of occurrence will allow a better estimate of the probable distribution of metrical structures in actual speech samples.

3. Frequency of occurrence

The MRC Database includes the frequency of occurrence statistics of Kucera & Francis (1967). For the lexical word sample summarized in Table III, the mean frequencies of occurrence are given in Table IV.

TABLE IV. Mean frequency of occurrence (per million words) of lexical words by metrical characteristics, MRC Psycholinguistic Database

Monosyllables	39·35
Polysyllables with initial primary stress	6·91
Polysyllables with initial secondary stress	3·19
Polysyllables with weak initial syllable	6·00

It can be seen that monosyllabic words indeed occur far more frequently than any other type. Multiplying out the statistics of Tables III and IV suggests that roughly 45% of lexical words encountered in average speech situations will be monosyllabic. A further 39% will be polysyllabic words with strong initial syllables, giving an average total of 84% of lexical words having strong initial syllables. Only 16% of lexical words are likely to have weak initial syllables and hence not to have their initial boundaries identified by the Cutler & Norris (1988) segmentation strategy.

However these estimates are in no way precise. The Kucera & Francis (1967) frequency count was based on a written corpus of American English. Accordingly we tested the preliminary estimates against the metrical characteristics of an actual corpus of spoken English.

4. A corpus of English conversation

The *London-Lund Corpus of English Conversation* (Svartvik & Quirk, 1980) consists of 34 samples, each consisting of between 5000 and 6000 words, of spontaneous conversation between educated adult native speakers of British English. The majority of the speech material comes from speakers who were unaware that they were being recorded. A frequency count of this corpus has been carried out by Brown (1984). This frequency count was analyzed in the same manner as the databases described above.

Many proper names in the corpus have been changed to preserve anonymity. Because it cannot be determined which proper names are not part of the originally spoken material, the Brown (1984) count omits all proper names. The remaining corpus consisted of 187 833 tokens (8985 separate words). From this a small additional number of non-words was omitted (e.g. *arrh*, *eigh*), leaving 187 699 tokens and 8933 separate words.

The corpus was classified by word class and by metrical structure, using the grammatical and phonetic information in the MRC Database; 1305 words, of which there were no phonetic transcriptions in the MRC Database, were analysed by hand.

In this corpus, 59% of the tokens were grammatical, or closed-class items: 110 736 tokens, a total which was achieved, however, by only 281 separate words. The metrical structure of the 76 963 lexical tokens is shown in Table V. It can be seen that the rate of occurrence of lexical words with strong initial syllables in this sample of real speech exceeds the rough estimates based on average written frequencies: 90% of the lexical tokens have strong initial syllables.

Although we can compute the total number of syllables in this corpus (or rather, in the 187 699 words we analysed), it is not possible to know exactly the total frequencies of

TABLE V. Metrical characteristics of initial syllables of lexical tokens in the *London-Lund Corpus of English Conversation*

	Sum	Proportion
Monosyllables	45 718	0·594
Polysyllables with initial primary stress	21 706	0·282
Polysyllables with initial secondary stress	1993	0·026
Polysyllables with weak initial syllable	7546	0·098

TABLE VI. Potential metrical characteristics of initial syllables of grammatical tokens in the *London-Lund Corpus of English Conversation*

	Potentially weak	Necessarily strong
Monosyllabic	95 316	8804
Polysyllabic	4732	1884

TABLE VII. Probable metrical characteristics of syllables by word class, in the *London-Lund Corpus of English Conversation*

		Open-class words	Closed-class words	Total
Word-initial syllables	Strong	69 417	10 688	80 105
	Weak	7546	100 048	107 594
Non-word-initial syllables	Strong	10 924	2960	13 884
	Weak	34 027	4267	38 294
Total		121 914	117 963	23 9877

strong and weak syllables. Many of the closed-class words, for instance, could potentially be spoken with either full or reduced vowels (*their*, *in*, *could* etc.); however, the transcript of the corpus is orthographic, not phonetic, and therefore does not resolve this ambiguity. Table VI presents the distribution of closed-class tokens by potential metrical structure. Over 86% of the closed-class items are monosyllabic, and nearly all of these are potentially weak; most of the polysyllabic words also have potentially weak initial syllables. Only 9·5% of the tokens were necessarily spoken with strong initial syllables (e.g. *ours*, *those*, *either*, *mightn't*). (The polysyllabic initially-strong class is not subdivided, since only four words had initial secondary stress.)

It is probably safe to assume, though, that by far the majority of the closed-class tokens (including most of the 6833 occurrences of *the*, the 5453 occurrences of *and*, and the 5006 occurrences of *a*) were in fact metrically weak. If all potentially weak closed-class monosyllables were indeed realized as weak syllables, then the distribution of syllables in the 187 699 analysed words would be as given in Table VII.

If the segmentation strategy proposed by Cutler & Norris (1988) were applied to this corpus, it would successfully locate the initial boundaries of 90% of the lexical items. The false alarm rate would also be low since 74% of strong syllables are indeed initial syllables of lexical items; only 11% are initial syllables of closed-class words, and 15% are non-word-initial syllables. In contrast, of the weak syllables 69% are initial syllables of closed-class words, and the majority of the remainder (26%) are non-word-initial. Only 5% are initial syllables of lexical words.

5. Towards an implementation of the strong-syllable segmentation strategy

Implementation of any component of a speech recognition system obviously involves many questions concerning how that component interacts with the system architecture

in general; if such questions are set aside for the time being, however, some characteristics of a strong-syllable segmentation algorithm may be tentatively proposed:

(1) *1.1 Lexical Structure.* The main lexicon contains only open-class (lexical) words; closed-class (grammatical) words constitute a separate list.
1.2 Segmentation. An initial segmentation process scans the input and places markers at the onset of each strong syllable.
1.3 Lexical Access.
1.3.1 If the initial string of the current input is not preceded by a marker, it is submitted to the closed-class list. If there is a marker, the string is submitted to the main lexicon.
1.3.2 The lookup process in both the main lexicon and the closed-class list returns the longest candidate consistent with the input, *except that* the occurrence of a marker indicating the beginning of a strong syllable will terminate the current lookup process, returning the longest word found so far (if any), and initiate a new lookup process in the main lexicon. When a match is found, the process returns to the input (i.e. starts again at *1.3.1*).
1.3.3 Failure of either lookup process leads to the same string being submitted to the other store.
1.3.4 Failure in both stores leads to backtracking, i.e. cancellation of a previous decision. This may involve accepting a shorter candidate word, if there is one; or undoing a word assignment for the preceding syllable and attaching it to the current input string; or over-riding a following marker and initiating a lookup process for a string with a medial strong syllable. (We have no evidence on which to base a proposal for ordering of these choices.)

The performance of this algorithm on the London-Lund conversational corpus could only be calculated by considering all words *in context*, which would be a very onerous undertaking. Accordingly we constructed a small corpus specifically for the purpose of testing the algorithm's operation. A native speaker of standard southern British English, unaware of the purpose of the task, read onto tape the "Rainbow Passage". Three prosodically trained listeners transcribed the metrical prosody of her recording. Her production is reproduced in (2); syllables which one or more of the transcribers marked as strong are preceded by a slash mark:

(2) When the /sun/light /strikes /rain/drops in the /air, they /act /like a /prism and /form a /rain/bow. The /rain/bow is a di/vision of /white /light into /many /beautiful /colours. /These /take the /shape of a /long /round /arch, with its /path /high a/bove, and its /two /ends ap/parently be/yond the ho/rizon. There /is, ac/cording to /legend, a /boiling /pot of /gold at /one /end. /People /look, but /no one /ever /finds /it. /When a /man /looks for /some/thing be/yond his /reach, his /friends /say /he is /looking for the /pot of /gold at the /end of the /rain/bow.

The algorithm as sketched in (1) is both extremely crude and a very stringent implementation of the proposed strategy. As we shall suggest below, there are some obvious ways in which its operation could and probably should be refined. Nevertheless, despite this lack of sophistication, and irrespective of which choices are made for further specification of the error correction component, the algorithm performs remarkably well on this passage. Of the 97 words in the text, 80 (82%) are initially assigned to the correct

store. Most of the initial misassignments, however, involve closed-class items; of the lexical items, 92% are correctly assigned to the main lexicon on the first pass. A further four strong syllables which are closed-class words (*these*, *is*, *it*, and *when*) will be correctly identified on the second pass (*1.3.3*). Backtracking will be required in a maximum of twenty cases (note that all judgements are based on the speaker's particular dialect):

(a) The initial syllable of *division*, containing a short [ɪ], will find no match in either store, so that the marker preceding its second syllable will have to be over-ridden.
(b) In two cases a strong syllable will find an incorrect match in the main lexicon: *like* will be recognized as a verb, *many* as *men*.
(c) In three cases an initial weak syllable will find an incorrect match in the closed-class list: *apparently* will be matched to *a*, and backtracking will occur when the only available entries beginning with the following syllables (*paranoid*, *paronoia* and *parenthetical*) fail to match the input; similarly, *according* will match *a*, and the point at which backtracking will occur will depend on whether the lexicon accepts *cording/chording* as a possible word; *horizon* will initially return *her*.
(d) The two instances of *beyond* will have the initial word assignments (*be*) undone when both stores fail to produced matches beginning *yond*.
(e) In six cases, the longest-alternative-first strategy will incorrectly attach a weak syllable to a preceding strong one: *shape of* will initially be matched to *shaper*, *high above* to *higher*, *pot of* to *potter* (twice), *reach his* to *reaches* and *he is* to *hears*. All except the last two of these will be immediately undone when no match is found to the following string.
(f) Finally, unnecessary lexical access attempts will be initiated in the middle of *sunlight*, *raindrops*, *rainbow* (three times) and *something*, leading to these compounds being parsed as two words. Except in the case of *rainbow* (where the meaning of the compound is not fully predictable from the meaning of the component words), this is presumably not serious.

Some of these misparses will require higher-level rejection to initiate backtracking: *rain bow*, *hears*, *reaches*, *a cording*, and *her* in *horizon*, for instance. But note that long sequences of backtracking will not be necessary: although *horizon* could potentially return *her rise on* or similar, a higher level analysis can reject the misparse from the first word since *her* is not a possible continuation after *the*.

It is interesting that the ways in which the postulated algorithm fails are precisely in line with further evidence from psycholinguistic studies of human auditory language processing. Firstly, misparses overwhelmingly involve incorrect omission of boundaries prior to weak syllables (e.g. *potter*), and incorrect postulation of boundaries prior to strong syllables (e.g. *a (p)parently*). Slips of the ear by human listeners show precisely the same pattern (Cutler & Butterfield, in prep.). Secondly, non-word-initial strong syllables trigger lexical access attempts (e.g. *rain bow*). Recent psycholinguistic studies have shown that non-word-initial strong syllables which are themselves words momentarily produce semantic activation; hearing *trombone* triggers lexical access of *bone* (Shillcock, in press).

Our proposal clearly begs many important questions. For instance, it assumes a reliable method of identifying strong syllables. We suggest that this will require at the very least a sophisticated simultaneous consideration of vowel spectral quality and relative duration (for which some estimation of rate of speech will be required). It also

assumes that the initial boundaries of strong syllables can be clearly identified, i.e. that there will be phonetic cues which will prevent placement of markers at, say, */darch* (in *round arch*) or */spath* (in *its path*). On the other hand, we have deliberately ignored certain sources of information which would almost certainly enable the algorithm's performance to be improved. For instance, prosody should prevent word boundaries which are also syntactic boundaries being overlooked; thus *reach his* in (2) should in fact never be parsed as *reaches* because the speaker clearly realised an intonational fall on *reach* to signal the end of an intonational unit. Although we were concerned to see how well the proposed strategy would perform in a very stringent form, it is clear that there are some obvious ways in which its operation could be substantially refined.

Finally, we should point out that even more radical proposals about the lexical access component of our suggested strategy exist in the psycholinguistic literature. For instance, several authors (Cutler, 1976; Bradley, 1980; Grosjean & Gee, 1987) have proposed that lexical words with weak initial syllables may be recognized via their strong syllables. Application of this proposal to the passage in (2) would involve, for example, the second syllable of *horizon* returning *horizon* as one potential candidate (along with *rise*, *riser* etc.) to be matched against the input string. Grosjean & Gee (1987) have extended this proposal to argue that *only* strong syllables are input to any lookup process; weak syllables provide so little phonetic information that they can only be evaluated via a pattern-matching process which is qualitatively different from the processing of strong syllables. As a corollary, however, closed-class words which can be strong syllables are represented in the main lexicon; thus in (2) *these*, *it*, *is*, *when*, *many* and *like* could on this proposal have been recognized on the first pass. On a somewhat different tack, Taft, Hambly & Kinoshita (1986) have suggested that recognition of prefixed words may involve the prefixes first being recognized as such, and lexical access being initiated via the stem. Since many weak initial syllables of lexical words are in fact prefixes, this proposal would have an effect not dissimilar to the preceding one. But it also implies that there is a stored list of possible prefixes. In (2) the word *division*, for example, might be recognized more efficiently by incorporation of the Taft *et al.* (1986) proposal if (a) the initial syllable found a match in the closed-class list indicating that it was a prefix, which in turn triggered (b) a lexical access process from *vis-* which considered *only* candidates in which *vis-* was non-word-initial.

6. Conclusion

The distribution of word types within the English vocabulary, combined with relative frequency of occurrence across types, provides an adequate basis for the successful implementation of a strategy of speech segmentation whereby strong syllables are assumed to be the onsets of lexical words. The strategy will result in few lexical word onsets being missed, and the false alarm rate will be relatively low. There are some obvious ways in which the strategy could be improved by consideration of additional sources of information; however, it is clear that even the unmodified strategy is well adapted to the characteristics of the English lexical vocabulary.

This research was supported by a grant (MMI 069) from the Alvey Directorate to Cambridge University, the Medical Research Council and STC Technology Limited. We thank Gordon Brown for making available his frequency count of the *London-Lund Corpus*, Uli Frauenfelder for making available the data on which Table II is based, Bill Barry and Charlie Hoequist for assisting with the prosodic analysis, and two anonymous reviewers for helpful comments on the paper.

David M. Carter is now at SRI International Cambridge Computer Science Research Centre, 23 Miller's Yard, Cambridge CB2 1RQ, U.K.

References

Bradley, D. (1978). *Computational Distinctions of Vocabulary Type*. Ph.D. thesis, unpublished. MIT.

Bradley, D. (1980). Lexical representation of derivational relation. In *Juncture* (M. Aronoff & M.-L. Kean eds). Saratoga, California: Anma Libri.

Brown, G. D. A. (1984). A frequency count of 190,000 words in the *London-Lund Corpus of English Conversation*. *Behaviour Research Methods, Instrumentation and Computers* **16,** 502–532.

Carlson, R., Elenius, K., Granstrom, B. & Hunnicutt, S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. *Speech Transmission Laboratory (Stockholm): Quarterly Progress and Status Report* **1,** 63–94.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* **33A,** 497–505.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics* **20,** 55–60.

Cutler, A. & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*. **14,** 113–121.

Friederici, A. D. & Schoenle, P. W. (1980). Computational dissociation of two vocabulary types: Evidence from aphasia. *Neuropsychologia* **18,** 11–20.

Grosjean, F. & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition* **25,** 135–155.

Jones, D. (1963). *A Pronouncing Dictionary of the English Language*, 12th edn. London: Dent.

Kucera, H. & Francis, W. N. (1962). *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

Shillcock, R. C. (in press). Speech segmentation and the generation of lexical hypotheses. *Cognition*.

Svartvik, J. & Quirk, R. (1980). *A Corpus of English Conversation*. Lund: Gleerup.

Taft, M., Hambly, G. & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology* **38A,** 351–366.